Contemporary Educational Psychology 73 (2023) 102160

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback

Yong Wu [a,*], Christian D. Schunn [b]

[a] School of Humanities, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Learning Research and Development Center, University of Pittsburgh, 3420 Forbes Ave, Pittsburgh, PA 15260, USA

ARTICLE INFO

Keywords:
Peer feedback
Explanations
Suggestions
Revisions
Learning

ABSTRACT

To deepen understanding of learning through peer feedback, the current study investigated the relationships between different peer feedback activities (organized into constructive vs active activities) and learning (i.e., transfer to new tasks), examining the nature of activities within provided feedback, received feedback, and revisions in response to feedback. Across five US high schools, 367 students in Advanced Placement classes participated, implementing common assignments and peer assessment rubrics. Provided/received comments and revisions in one assignment, and writing improvements observed in a second assignment were exhaustively coded and subjected to hierarchical model regression analyses. Results showed that constructive activities (providing explanations and making revisions after receiving explanations or providing suggestions) were consistently associated with learning, whereas passive (e.g., receiving feedback without making revisions) or active activities (e.g., implementing specific suggestions) were not. Further, the effects of received feedback on learning were mediated by the number of revisions. Theoretical and practical implications of the findings are discussed.

1. Introduction

Feedback has long been considered a central part of teaching and learning (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Liu & Carless, 2006; Topping, 1998). However, accounts of feedback (e.g., teacher or automated feedback) often involves conceptualizing the learner as a *recipient* of useful information, a passive learning mode. By contrast, more recent theories of learning have placed greater emphasis on active, constructive, and interactive processing as especially useful for learning (Chi et al., 2018; Chi & Wylie, 2014). Peer feedback is an interesting case of growing importance in classroom instruction and large-scale online teaching (Liu & Carless, 2006; Wu & Schunn, 2020a). However, much research has focused either on the passive reception of peer feedback (e.g., is it as helpful as teacher feedback?; what kinds of received feedback is more persuasive?) (Leijen, 2017; Nelson & Schunn, 2009) or conceptualized it in simple binary terms (giving vs receiving) (Huisman et al., 2018; Lundstrom & Baker, 2009; Wu & Schunn, 2020b). In reality, there are many elements to peer feedback (reading peer's documents, evaluating documents in light of criteria, constructing advice to peers, receiving evaluations, interpreting feedback, revising in response to feedback provided and received), and several of these elements require students to act as more than passive recipients of feedback. Thus, peer feedback is inherently less passive than teacher feedback because in peer feedback students need to evaluate others' work as assessors in addition to receiving feedback as assessees. Most importantly, we argue here that these different peer feedback elements are varied in their learning affordances, and we report the results of an empirical investigation based on the ICAP framework (i.e., Interactive, Constructive, Active, and Passive framework) that tests the learning effects of different peer feedback elements in terms of this new theoretical analysis.

The main aim of the present study is to investigate the relationships of three types of overt learning activities found within peer feedback to learning. Although grounded in underlying psychological processes, the direct empirical focus is on specific overt learning activities because teachers can directly control such activities, they are easy to unambiguously define, and they are easier to empirically study (Chi, 2009). Therefore, the findings provide teachers with important insights for designing and implementing peer feedback activities to maximize learning outcomes. The three types of learning activities examined here are constructive activities (i.e., revisions after receiving explanations, providing explanations/suggestions), active activities (i.e., revisions

* Corresponding author.
E-mail addresses: wuyong@bupt.edu.cn (Y. Wu), schunn@pitt.edu (C.D. Schunn).

https://doi.org/10.1016/j.cedpsych.2023.102160

Available online 26 January 2023

after receiving suggestions), and passive activities (i.e., receiving explanations and suggestions without making revisions). Interactive activities did not occur in the study, but they are possible in other configurations of peer feedback and are therefore considered in the discussion.

2. Theoretical background

2.1. The ICAP framework

The Interactive, Constructive, Active, and Passive (ICAP) framework characterizes learning opportunities in terms of overt learning activities and is used here as a theoretical lens through which we explore how students learn to write from peer feedback. Chi and her collaborators differentiate four levels of overt learner activities that have been shown to produce different learning gains: interactive > constructive > active > passive (Chi, 2009; Chi & Wylie, 2014; Chi et al., 2018).

Passive learning, the lowest level, occurs when learners receive information without doing anything observable related to learning. The typical examples of this type include listening to a lecture or reading a passage without doing anything else (e.g., without active notetaking).

Active learning involves learners who are physically moving, manipulating some parts of the learning materials. Examples of active learning activities are taking verbatim notes or highlighting important information while listening to a lecture or reading a passage. The physical activity may prevent learners from zoning out during instruction. Additionally, active learning may not only lead learners to focus their attention on specific information (e.g., highlighted information when highlighting text while reading), but may also result in more rehearsal or practice.

Constructive learning involves learners generating new knowledge or information beyond what they are receiving from instruction, often by making connections either across information that was received or by making connections between information that was received and prior knowledge. Typical examples of constructive activities include students taking notes by summarizing or explaining what they are learning using their own language rather than simply copying presented information.

Interactive activities involve constructive utterances with turn taking between learners. In other words, learners share information beyond what is presented in the learning materials (i.e., they are constructive), and they substantially contribute to a discussion of this information. As a result, participants can learn new ideas, correct incorrect understandings, and can reach a deeper understanding of the learning materials. Typical examples of interactive activities include arguing for a position or asking-and-answering questions in pairs.

Prior research using ICAP has been used to classify traditional classroom activities given to students (e.g., Wiggins et al., 2017), teacher behaviors (e.g., Chase et al., 2019), or focused on online learning (e.g., Raković et al., 2020; Wekerle et al., 2020). For example, Raković et al. (2020) analyzed ten characteristics of students' online posts, and found three characteristics (i.e., disagreeing, comparing, and making claims) predicted writing performance. Wekerle et al. (2020) found the strongest learning outcomes for interactive digital learning activities and no learning benefit for passive learning activities. Webb et al. (2021) found that explaining one's own ideas and engaging with others' suggestions are productive for math learning, especially for less proficient students in math. The ICAP framework has not previously been used in peer feedback research, even though the framework seems particularly relevant to understanding peer feedback.

2.2. Prior research on peer feedback activities

2.2.1. Studies of receiving and providing feedback

Peer review consists of peer assessment (the scoring of documents) and peer feedback (the comments, written or oral, provided to peers on their documents). Here we focus on peer feedback because it is theoretically and empirically better connected with learning outcomes and it enables students to learn about their strengths and weaknesses, and how to improve their work (Liu & Carless, 2006; Lu & Law, 2012).

Peer feedback has received a lot of attention in both L1 (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Topping, 1998; Wu & Schunn, 2022) and EFL/ESL writing instruction (Leijen, 2017; Lundstrom & Baker, 2009; Wu & Schunn, 2019). The majority of this literature has focused on the effects of receiving feedback on writing, which we consider here as a passive activity when considered alone. Some studies of receiving peer feedback on its own (i.e., received but not actioned upon) find that it tends not to improve learning outcomes (i.e., performance on a new writing task; Wu & Schunn, 2020b). However, a number of studies have tracked the effects of received feedback into revisions, finding that receiving peer feedback does lead to revisions and improvements in quality of revised drafts (Huisman et al., 2018; Nelson & Schunn, 2009; Patchan et al., 2016; Wu & Schunn, 2020a).

A few studies have examined the effects of providing feedback on learning (i.e., assessed on a new writing task). Lundstrom and Baker (2009) divided students into assessees (received feedback from peers, did not review, and used feedback to revise papers) and assessors (reviewed, did not receive feedback, did not revise papers). They found that assessors made more significant gains on posttest writing than did assessees. However, those who received feedback *and revised* on the basis of the received feedback did also improve, supporting the learning effects of receiving feedback when accompanied by revising. Prior research comparing a full reviewing condition, a reading-only condition, and a no-treatment control found that reviewing was associated with improvements in writing a new essay (Cho & MacArthur, 2011; Philippakos & MacArthur, 2016).

Studies have also examined the effects of providing feedback on revising, finding that students do more revising, producing improvements in essay quality after providing feedback to their peers (Cho & Cho, 2011; Huisman et al., 2018; Lu & Law, 2012). Further, the amount of revision that occurs increases when the provided feedback includes suggestions for how to revise (Lu & Law, 2012) or explanations (Huisman et al., 2018).

2.2.2. Studies of explanations and suggestions and learning to write

Peer feedback can include specific features (e.g., mitigation, localization, praise, explanations, suggestions), but not all features increase the likelihood of revisions or learning. For example, the number of praise comments received is not related to (Cho & MacArthur, 2011) or is negatively related to later writing quality (Cho & Cho, 2011). Narciss (2008) classified feedback types into 1) simple evaluative feedback (e.g., whether the task has been done correctly or incorrectly) and 2) elaborated feedback that includes extra information (e.g., providing explanation, suggestions). Elaborated feedback had a more positive effect on writing performance than did simple evaluative feedback (Hattie & Timperley, 2007; Narciss, 2008). For example, assessees are more likely to revise when they receive suggestions for how to improve (Leijen, 2017; Nelson & Schunn, 2009) or explanations of the problems or suggestions (Gielen et al., 2010; Huisman et al., 2018; Strijbos et al., 2010; Wu & Schunn, 2020a,c).

Explanations are regularly argued to be an essential component of feedback that is especially likely to influence whether students follow the feedback in their revisions. Receiving explanations can enhance cognitive processes of feedback (Bolzer et al., 2014), and a lack of explanations in received feedback might demand too much effort for students to figure out how to integrate peer feedback into the essay (Bolzer et al., 2014). Some research (e.g., Gielen et al., 2010; Huisman et al., 2018; Wu & Schunn, 2020a,c) found a positive effect for received explanations, while two studies (Nelson & Schunn, 2009; Tseng & Tsai, 2007) reported the opposite. The authors of these two studies argued that receiving explanatory peer feedback may have hurt student writing performance because novice writers were not be able to provide high-quality explanations (Nelson & Schunn, 2009; Tseng & Tsai, 2007).

Students do not benefit from receiving explanatory feedback when the explanations are of poor quality (Lu & Law, 2012).

Providing explanatory feedback may also influence assessors' own writing. Wooley et al. (2008) found that giving explanations was positively associated with assessors' subsequent writing quality. Further, Cho and MacArthur (2011, p. 78) noted "It [giving explanations] has larger effects than receiving explanations". Providing explanations also is similar to self-explanation, which has long been argued as useful for learning (Chi et al., 1994).

Receiving feedback with constructive suggestions on how to improve work can be helpful because students may not know how to revise (Price et al., 2011). Multiple studies have shown that feedback including suggestions resulted in stronger revisions than when the feedback just indicated that a response was wrong (Bangert-Drowns et al., 1991; Nelson & Schunn, 2009; Tseng & Tsai, 2007).

Providing suggestions has also been found to be positively related to assessors' revised work (Lu & Law, 2012) or new writing (Cho & MacArthur, 2011), perhaps because through coming up with suggestions, students generate new knowledge (i.e., how to solve new problems) and learn constructively.

2.3. The current study

In the current study, we test an ICAP-based theoretical analysis of peer feedback activities and thereby address three key gaps in the literature. First, the study is conducted in the context of reciprocal peer review where students both receive and provide peer feedback, a research design with ecological validity. Few studies of peer feedback have tested the separate effects of providing vs receiving peer feedback (Double et al., 2020; Wu & Schunn, 2021). Second, feedback with different features is treated separately in the current study. Prior research has typically analyzed the learning benefits of received feedback and provided feedback in global terms (e.g., Huisman et al., 2018; Wu & Schunn, 2020b), and therefore the larger effects of provided or received feedback with specific features might have been masked (Gielen et al., 2010). Third, we connect specific peer feedback activities to both revision (i.e., later performance within the same writing task) and learning (i.e., performance on a new writing task). Most prior research only investigated whether specific feedback activities resulted in revisions or improved drafts (e.g., Lu & Law, 2012; Tseng & Tsai, 2007). Even when a writer improves a piece of writing upon receiving feedback, they may not have learned a concept or mastered a skill such that it can be applied to a new writing task (DiPardo & Freedman, 1988; Hu & Lam, 2010; Patchan & Schunn, 2015).

We also examine the effects of peer feedback activities separately for high-level (e.g., meaning, argument, organization) and low-level (e.g., spelling, grammar) aspects of writing. Prior research has observed differences in effects on student performance for high-level vs low-level aspects of writing. For example, students often make more learning gains through peer feedback on high-level aspects of writing (Cho & Cho, 2011; Lundstrom & Baker, 2009; Wu & Schunn, 2020b). In addition, high-level issues may require different supports for learning than low-level aspects of writing. For example, Liou and Peng (2009) found that high-level problems needed more explanatory information.

The ICAP framework emphasizes overt learning behaviors: behaviors that can be observed and explicitly reflect students' likely engagement in learning (Chi & Wylie, 2014). Thus, learning activities (like different kinds of activities or ways of completing an activity within peer feedback) can be differentiated by the kinds of overt behaviors that each activity requires, which are categorized by the cognitive processing they directly entail. For example, receiving feedback without requiring students to highlight parts of the feedback or making any revisions would be characterized as passive learning because no actions are required. Students might spontaneously self-explain their received feedback, but this would likely be uncommon because most students rarely self-explain during reading and even direct instructions to self-explain are often ineffective in practice (Rittle-Johnson et al., 2017; Patchan & Schunn, 2016), particularly without training on self-explanation (McNamara, 2004). Similarly, providing feedback would be characterized as constructive learning because students must generate feedback comments rather than just reading essays. Note that ICAP does not consider cognition and behaviors that students might choose to add on the own (e.g., choosing to self-explain information that is read, or choosing to discuss a lecture afterwards with a peer or family member). Not only are those difficult to observe, but they also are likely rare (Chi & Wylie, 2014).

2.4. Research questions and hypotheses

We examine a range of typically-occurring peer review-related activities (providing, receiving, and revision) in terms of their relative impacts on learning, teasing apart the relative benefits to particular kinds of comments (i.e., suggestions and explanations) received and provided on learning outcomes. Fig. 1 provides a summary, focusing on the specific peer feedback elements that will be tested.

Receiving feedback tends to be at the lower ICAP levels, but can still vary depending upon what students do with the feedback. Reading received peer feedback without making revisions, *per se*, is a clear example of passive learning, regardless of the quality or characteristics of feedback that is received (see upper left of Fig. 1). Received comments typically indicate that some aspect of the information assessees had previously stored is incorrect. However, the newly received information (i.e., the comment) is likely to become inert knowledge if it is not used (e.g., in a reflection activity, revision plan, or revision behavior), and therefore is likely to only be accessed when it is specifically activated. If students highlight parts of the received feedback (not shown on the figure, which focuses on typically observed peer feedback activities), this is active learning. When learners receive specific suggestions for how to revise and then directly implement the suggestions in their revisions without considering other possible revisions, this is also a kind of active learning (see upper right of Fig. 1). Internally, active activities such as implementing received suggestions in revisions can lead the assessee to strengthen knowledge through practice or update their knowledge relevant to the identified problem.

Different from passive learning that tends to store information in isolation and active learning which may strengthen knowledge or update incomplete schema, constructive learning focuses on general inferences creating new conceptual knowledge and justifications for procedural knowledge (Chi & Wylie, 2014). We argue that constructive learning can be said to occur when students provide feedback to their peers, either in the form of suggestions for how to improve their documents or in the form of an explanation of the problem or of the suggestion (see lower left of Fig. 1). However, we also note that constructive learning can also be involved during revising in response to peer feedback, particularly when the feedback leads to revision but did not fully describe the necessary revision. Explanations received from peers without a specific suggestion are a clear case of feedback requiring construction by the assessee in order to produce a revision (see upper right of Fig. 1). Further, providing feedback (explanations or suggestion) can also motivate revision in the assessor's own documents, and this connection is again a form of constructive learning because some adaptations or ideas beyond the given information is likely required in order to translate problems or solutions that are noted in a peer's document to the assessor's own document. Here providing feedback with explanations may be particularly helpful in guiding the translation into revisions in the assessor's own document (see connection from lower left to upper right of Fig. 1).

The theoretical analysis using ICAP as summarized in Fig. 1 suggests that *receiving* feedback will rarely lead to learning if it does not lead to revisions, whereas *providing* feedback will regularly lead to learning, because it involves both a higher quality direct learning opportunity as well as having a mediated pathway through revisions. We organize our

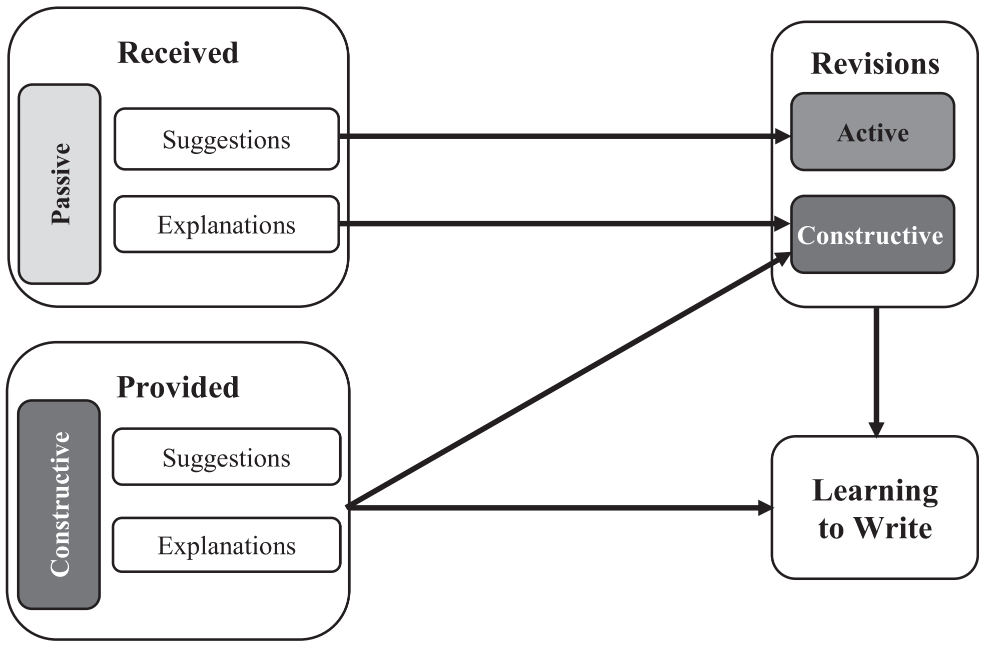

Fig. 1. Empirical tested feedback-to-learning model. Receiving peer feedback (suggestions or explanations) without making revisions, *per se*, is a passive learning process that is therefore unlikely to directly lead to significant learning. Receiving specific suggestions for how to revise and then directly implementing the suggestions in revisions is a kind of active learning. Making revisions based on received explanations without a specific suggestion is constructive learning. Providing suggestions and explanations is also constructive learning and can involve additional constructive learning by leading students to make revisions in their own documents.

analyses around the following three research questions. Note that research question 1 is about predicting revisions rather than learning, the foundational indirect pathway from reviewing to learning via revisions, to establish the first step in the mediation model that is formally tested in research question 3. Thus, the components of that research question are not a direct test of ICAP since they are about performance (i.e., revising), not learning outcomes *per se*. Also, note that the hypothesis focuses on observables (e.g., number of received explanations, number of provided suggestions, number of revisions).

1 What types of peer feedback activities are associated with revisions? We assumed that:

H1a: Number of received explanations predict number of revisions;

H1b: Number of received suggestions predict number of revisions;

H1c: Number of provided explanations predict number of revisions;

H1d: Number of provided suggestions predict number of revisions.

2. What types of peer feedback activities are directly associated with learning? We assumed that only active and constructive activities would have substantial direct associations with learning:

H2a: Students learning is predicted by number of revisions;

H2b: Students learning is predicted by number of provided explanations;

H2c: Students learning is predicted by number of provided suggestions.

3 What are the relative contributions of different peer feedback activities to learning through revisions or directly? According to ICAP, we predicted:

H3a: Weakest associations with received suggestions (one indirect path, only active learning);

H3b: Weak associations with received explanations (one indirect path, constructive learning);

H3c: Stronger associations with provided suggestions and explanations (direct and indirect paths, both constructive).

3. Method

3.1. Participants

The participating students came from five high schools. These schools participated in a larger project studying the use of different versions of an online peer assessment system. The sample consisted of 367 students (58% female, 42% male; 51% Caucasian, 20% Asian, 5% Hispanic/Latino, 5% African American, and 19% no race or ethnicity reported). The student ages ranged from 16 to 19 years (*Mean* = 17.2, *SD* = 0.5). Forty-nine percent of the participants came from Title I schools, which receive extra federal financial support because they have high percentages of students from low-income families (US DOE, 2018) . The remaining participants were from non-Title I schools (i.e., serving predominantly middle- and high-income families).

All students were enrolled in Advanced Placement (AP) Language and Composition, a high school course that is meant to be equivalent to first-year university writing courses. Over a half-million students took the AP Language & Composition exams in the 2017–18 school year (College Board, 2018). Many students struggle with this course; the mean score is just below a 3, which is the lowest value any university would accept as equivalent to passing their own course. The writing challenges coupled with the large volume of students participating in this course call for more research to be conducted in this context.

Students from a given school were taught by a single instructor across multiple sections. The schools were located in five different states (i.e., Kentucky, New York, California, Delaware, and Texas), and the instructors had taught English and the AP writing course for multiple years. As a pre-condition for participation, teachers had to have laptops in the classroom or had access to a computer lab to use the online peer assessment system, agreed to participate in an "act-as-a-student" online training exercise, and agreed to implement two shared writing tasks as designed using a shared peer assessment system at a similar time of year. The instructors also agreed to use provided scripts (e.g., how to introduce the system and the peer feedback process) in class to have a standardized invention across classrooms and schools. The instructors had no knowledge of the specific hypotheses of the current study.

3.2. Materials

Peer assessment tool. The online peer review system, Peerceptiv (Schunn, 2016), is an online peer assessment system developed to address the need for large scale writing assessments, and it is being used by a large number of students in high schools, colleges, and universities throughout the world. With the system, teachers can assign writing tasks in large class settings and maintain control over the peer feedback process. Students can receive feedback from multiple reviewers and provide feedback based upon rubrics specified by the teacher; in the current study, all teachers used the same rubrics. Additional functions are included to encourage students to provide quality feedback, such as accountability measures, well-designed rubrics, and suggestions for including helpful information in feedback (Schunn, 2016). For example, students as assessees are required to evaluate the quality of the feedback they received (called back-evaluations). In the current study, these back-evaluations also provide evidence that assessees received the feedback their assessors submitted into the system. Indeed, 90% of students completed their back-evaluations, and thus we can generally rule out the possibility that students did not revise or learn from received feedback because they did not even read the feedback provided to them.

The peer assessment process consisted of separate phases. Instructors assigned the writing task, including assignment details such as essay topic, length, and submission deadline. Students submitted their essay to the web-based system, and it distributed the essay to four random peers at their school. Supported by the rubrics organized into dimensions of writing, assessors anonymously evaluated their peers' essays with qualitative and quantitative components (written comments and scores).

Writing tasks. A central focus in the AP course is an evidence-based, analytical writing task, which is challenging for secondary students (National Center for Education Statistics, 2012). Participants were asked to read a persuasive source essay and then write a one-page argument based upon an analysis of the source essay. In particular, students needed to describe what rhetorical strategies were used, and then analyze how the strategies were used to support the author's thesis, citing evidence from the source essay. For the current study, two such writing tasks were used (one for pre-assessment and learning opportunity; the other for post-assessment of learning), taken from previous years' exams, with the same requirements and identical rubrics. The only difference between the two writing tasks involved the source passages, which were of roughly similar length, complexity, and reading levels. The first was about the separation between people and nature due to the development of technology, and the second was about the effect of migrations. As a national and high-stakes exam, the writing tasks were developed by the College Board to be of similar difficulty. Further, expert grading of these writing tasks in a larger study with more participants across more schools found identical average scores when each was used as the first writing task (Schunn et al., 2016; Wu & Schunn, 2020a).

The reviewing and grading rubrics of these two writing tasks (see Appendix A) were adapted from the College Board expert scoring guide in order to be more student-friendly (Schunn et al., 2016). The high-level writing rubrics focus on thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, and organization. The low-level writing rubrics focus on control of language and conventions. The rubrics students used were specific to the evidence-based analytical writing task rather than generic writing rubrics disconnected from context and genre.

Genre-specific rubrics were used to orient students' attention to the specifics of the genre and develop their knowledge and skills related to this genre. Prior research has found positive effects of genre-specific rubrics on writing of assessors (Philippakos & MacArthur, 2016; Yu, 2021) and assessees (Wu & Schunn, 2020b), although it can be challenging for students to provide genre-specific feedback (Yu, 2021). Students received peer review training and were required to conduct peer review based upon the genre-specific rubrics.

3.3. Measures

Writing quality. Four trained writing experts with years of experience teaching writing rated all the essays (i.e., first drafts of two consecutive writing tasks) based on the same 1-to-7 scale rubrics used by the students (see Appendix A). Each rater gave each essay eight ratings, six ratings covering high-level dimensions of thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, organization, and two ratings covering low-level dimensions of control of language and conventions. Ratings across dimensions were averaged to produce a high-level score and low-level score for each essay. Each essay was rated by at least two raters with substantial reliability (*Kappa* = 0.75 for high-level scores and 0.72 for low-level scores; Anthony & Joanne, 2005). High-conflict cases (>1.5 difference) were discussed to resolution. For low-conflict cases, a mean score was used in analyses. Students' learning was operationalized as improvement in first-draft quality across the two consecutive writing tasks.

Amount of feedback provided and received by subtypes. Feedback was systematically coded to determine how many comments with explanations, suggestions, and praise comments each student received and provided. Although not the primary focus, amount of praise was included as a statistical control. What students learn from reading good essays may influence revisions (Lu & Law, 2012; Patchan et al., 2016; Tseng & Tsai, 2007) and performance on new writing tasks. Provided praise is an indicator of students noticing positive features in the documents they read. In the current data set, the quality of a draft judged by experts was positively related with the number of praise comments provided on that draft by a given assessor ($r = 0.32$, $p < 0.001$).

The first step was to segment the peer comments into independent units (*Kappa* = 0.93), defined by a focus on one aspect of writing. The reviewing interface required students to enter separate comments into different boxes, so most comments were already segmented. However, sometimes students combined different issues in one textbox. For example, "*Some of your points were not backed up with direct quotes. In your third paragraph, you talked about the end of the passage, but did not include any quotes or direct evidence to support your claims. Throughout the essay, there were multiple grammatical errors, which at times conflicted with the readers understanding.*" This comment included two idea units, with the first one (first two sentences) indicating a high-level problem and the second one (last sentence) on a low-level problem. Then, coders further segmented the comments including more than one problem when coding feedback scope and features.

Then the scope of each feedback comment was double-coded by trained coders: high-level or low-level (*Kappa* = 0.91). High-level feedback refers to the feedback on essay content and organization such as thesis, argument, rhetorical strategies, evidence for claims, explaining evidence, organization. Low-level feedback includes the feedback on control of language and conventions.

Then the comments were coded as summary comments, praise comments (i.e., pure praise comments), and negative evaluative comments by two trained coders (*Kappa* = 0.93). Only praise comments and negative evaluative comments were further analyzed because they have been found to influence learning outcomes (e.g., Patchan et al., 2016; Tseng & Tsai, 2007; Wu & Schunn, 2021).

All the negative evaluative comments were double-coded by four trained coders for the presence or absence of three major features: Explicit identification (*Kappa* = 0.84), Explanations (*Kappa* = 0.81), Suggestions (*Kappa* = 0.91). Note that explanations could have focused on either an identified problem or a suggestion, but in practice they were given almost exclusively for problems. A given negative evaluative comment could have between one and three of these features. However, the statistical analyses did not examine explicit identification. First, identification did not have a strong foundation for changing motivation the way that praise does or cognitive processing the ways that explanations and suggestions do (e.g., Bolzer et al., 2014; Gielen et al., 2010; Lu & Law, 2012; Tseng & Tsai, 2007; Wu & Schunn, 2020a). Second, although a large number of negative comments did not identify problems explicitly, they indicated problems in an implicit way (e.g., "*A better transition to your thesis statement would be something about how technology hinders a relationship with nature, as opposed to in the past*".). See Appendix B for coding definitions and examples of explanations, suggestions, and praise comments. Disagreements were resolved through discussion with a third coder. The coders did not know the purpose of the study.

Finally, the number of comments containing explanations, the number of comments containing suggestions, and the number of praise comments for each student were calculated. This was done once from the perspective of an assessee (e.g., total number of each that were received) and again from the perspective of an assessor (e.g., the total number of each that were provided). For example, if a student provided 6 explanations in comments in their first completed review, and then 7, 3, and 10 explanations in the remaining three reviews they completed, this would have been aggregated to be 26 total provided comments with explanations.

Amount of revisions. Students made revisions to the first document based upon the peer feedback received and provided, or based upon self-revisions or other sources. The MSWord Compare Documents function was utilized to identify revisions between the first and second drafts of the first writing task. Each revision was coded by at least two coders in terms of its content focus, high-level or low-level (*Kappa* = 0.71). If a revision changed the original meaning, it was a high-level revision (see examples in Appendix B), and otherwise it was coded as a low-level revision. Revisions in a document were aggregated to create two measures: number of high-level revisions and number of low-level revisions.

Control variables. In addition to the number of praise comments, additional variables that have been previously found to be correlated with students' writing performance were included as controls: students' first draft score (i.e., Task 1 score), School Type (Title I or not), and Gender (Female = 1, Male = 0) (see Appendix C). Task 1 score was included as a control variable in predicting revisions because students with high first draft scores had less incentive to revise. Students from different types of schools and students of different genders might perform differently in peer feedback and writing (Wu & Schunn, 2020a, c). Title I is a federal school type designation in the US frequently used in educational research. Schools with at least 40% of enrolled students coming from low-income families can use Title I funds to help students learn (US DOE, 2018), and Title I school students have been found to be in need of more assistance in writing (Applebee & Langer, 2011; Wu & Schunn, 2020b).

3.4. Procedure

At the start of the study, the teachers were trained in the use of the reviewing system, the shared peer-review rubrics, and how to introduce the system/process to their students. The teachers participated in an "act-as-a-student" online exercise, submitting essays to the system and providing feedback to others. After that, a follow-up online training on the teacher interface was provided, including walking through simple procedures for how to teach students use the system. Then teacher provided their students with an in-class training on how to use the system and how to provide feedback comments, consisting of two phases: demonstration and practice. For the demonstration, students read a sample essay and then were shown example feedback comments for the essay. Some example comments were helpful (e.g., explanatory and constructive), while others were not (e.g., general and vague). Then students discussed as a class what made feedback comments helpful. In the practice phase, students read the second sample essay and provided feedback on it with a partner based on the rubrics. The class as a whole held a discussion on the feedback and ratings that were generated. At this point, the class discussed the peer review rating scales and students received calibration feedback on their use.

After submitting first drafts to the system, assessors had one week to respond to the four peers' essays at their school using the given rubrics. They were required to provide at least one comment on each of eight dimensions and scored each dimension on a 7-point scale. The system distributed peer feedback to assessees. Assessees then revised their essays based on the peer feedback and submitted the revised draft to the system. After students finished the first writing task, they completed the second writing task after a one-week interval.

3.5. Data analysis

Three hundred and sixty-seven students from five schools were included in the analyses, representing 81% of the students who participated in the peer feedback process across the five schools. The research project initially focused on only two schools and exhaustively coded all of their data. To improve statistical power and generality, additional schools were added to allow for multiple Title I and non-Title I schools to be included. Because the time intensive nature of the coding process, not all students from the additional school were included. Since the main outcome variables depended upon access to the submitted draft, we excluded students who did not submit a first or second draft (as high as 18% for two of the schools). At the third additional school, there was a large number of students, and consideration from a related research project led us to select a subset of students (essentially a random half within each participating class in that school).

The two groups, the selected students and the excluded students, were compared in terms of their gender, race, school title, age, and task 1 scores. x^2 tests showed that the two groups were not significantly different in terms of their gender, race, and school title (see Appendix D). Independent t-tests indicated no significant differences between the two groups' ages and task 1 scores (see Appendix E).

Multilevel regression models including a random intercept for school were conducted using Stata 16 given the nested structure of the data (students nested within teachers/schools). Nesting of students within classrooms was not included because students conducted the peer feedback online and were placed in one large multi-classroom reviewing pool across all participating classes at that school (with one participating teacher per school). Instead, robust standard errors with clustering by teacher were used to adjust for variation at the teacher level. Indeed, the ICCs (intraclass correlation coefficient) for both high-level and low-level learning were 11%, indicating school differences were a factor and nested structure to the data should be addressed. The variables were grouped into student-level variables (i.e., # of received explanations/suggestions/praise comments, # of provided explanations/suggestions/praise comments, # of revisions, task scores, gender) and a school-level variable (i.e., school title). Discrete predictor variables (i.e., # of received explanations/suggestions/praise comments, # of provided explanations/suggestions/ praise comments, # of revisions) and the continuous variable (i.e., task 1 scores) were grand-mean centered.

Different types of regression analyses depending on outcomes were conducted to investigate the relationships between amount of feedback provided and received with specific features, amount of revisions, and learning. To investigate the associations between feedback variables to revisions in the same writing task (RQ 1), two sets of multilevel negative binomial regressions were conducted (see Models 1 and 3 in Table 2). Number of revisions are count variables, which are typically right skewed. Because variance and means were not equal (for high-level revisions, $x^2(1) = 658.1$, $p < 0.001$; for low-level revisions, $x^2(1) = 1416.4$, $p < 0.001$), negative binomial regression rather than Poisson regression was selected (Coxe et al., 2009; Hilbe, 2007). Incidence Rate Ratios (IRR) were presented as effect size estimates for these regressions (i.e., values < 1 for negative relationships, values = 1 for no relationship, and values > 1 for positive relationships).

Multilevel linear regression models were used when the outcome was Task 2 high-level or low-level scores (RQ 2) because those scores were normally distributed. To measure learning, Task 2 scores were used as the dependent variable and Task 1 scores were included as predictors. This approach of controlling for prior performance rather than calculating improvement scores was used because improvement scores often produce regression-to-the-mean artifacts in regression analyses. The data satisfied the homoscedasticity assumption based on the scatterplots of studentized residuals and unstandardized predicted values. There were only two outliers (one high-level or one low-level task 2 score) according to leverage, studentized residuals, and Cook's *D* statistics (Aguinis et al., 2013; Fox, 1991). These outliers were replaced with the next-highest scores (Strijbos et al., 2010); however, the results were identical when scores were left untransformed. Inspection of correlations among predictors and outcomes (see Appendices F and G) revealed that multiple predictors were correlated with each other and the outcomes, but no predictors were so highly correlated with one another that multicollinearity problems would likely occur. However, sensitivity analyses were performed to test the robustness of the primary findings concerning revisions and learning by comparing full models including all predictor variables and reduced models including independent variables of interest or only significant independent variables.

To formally test the mediation effects of feedback to learning via revisions (RQ 3), multilevel mediation analyses were conducted using the MLMED macro for SPSS 25 (Hayes & Rockwood, 2020). It employed 10,000 bootstrapped samples to estimate confidence intervals. If the confidence interval does not include zero, the indirect effect is significant. In each case, high-level and low-level aspects of writing were treated as separate datasets for analysis.

4. Results

Table 1 presented the means and *SD*s for all main variables in the analyses. As an indicator that measurable learning had occurred over this short time-scale, students on average improved their writing performance in Task 2 relative to Task 1: mean first draft ratings were higher for Task 2 than for Task 1 in both high-level scores ($t(366) = 5.29$, $p < 0.001$, $d = 0.28$), and low-level scores ($t(366) = 4.71$, $p < 0.001$, $d = 0.24$).

4.1. Relationship of feedback activities to revisions (RQ1)

Table 2 presented the specific models used to predict high and low-level revisions (see Models 1 and 3). Number of received explanations predicted the number of high-level revisions, partially confirming H1a because this relationship was not observed for low-level aspects of writing. Number of provided suggestions predicted the number of revisions for both high-level and low-level aspects of writing, confirming H1d. The effects were generally larger for low-level revisions. H1b was partially confirmed since number of received suggestions only predicted number of low-level revisions. H1c (i.e., number of provided explanations predict number of revisions) was not supported. None of the other control variables were significant predictors of revisions. Sensitivity analyses showed that these results were robust for high-level and low-level revisions across models that removed control variables and non-significant key predictors (see Tables H1 and H3 in Appendix H). Note that all the sensitivity analyses were conducted separately for high-level and low-level aspects of writing.

Table 1
Predictor and Outcome Variable Means and Standard Deviations.

Variable	High-level comments		Low-level comments	
	Mean	*SD*	*Mean*	*SD*
Predictors *Received*				
Explanation	6.57	5.46	1.17	1.10
Suggestion	10.43	6.41	1.85	1.43
Praise	7.98	5.35	1.25	1.09
Provided				
Explanation	6.91	5.90	1.27	1.40
Suggestion	11.20	7.78	2.04	1.78
Praise	7.75	5.11	1.22	1.17
Control				
School title	0.49	0.50	0.49	0.50
Task 1 score	4.62	0.80	4.94	0.84
Gender	0.58	0.50	0.58	0.50
Outcomes				
Revisions	4.99	4.92	5.98	6.95
Task 2 score	4.84	0.76	5.13	0.79

Table 2
Summary of Analyses Predicting Revisions and Learning (i.e., Task 2 Score Controlling for Task 1 Score).

Model	High-level aspects of writing		Low-level aspects of writing	
	Negative binomial regression: Revisions (Model 1)	Linear regression: Learning (Model 2)	Negative binomial regression: Revisions (Model 3)	Linear regression: Learning (Model 4)
Measure of effect size	IRR (SE) [95 % CI]	Coefficient (SE) [95 % CI]	IRR (SE) [95 % CI]	Coefficient (SE) [95 % CI]
Predictors				
# Received				
Explanations	**1.042*** (0.01) [1.02, 1.06]** *p = 0.000*	−0.009 (0.01) [-0.03, 0.01] *p = 0.25*	1.112 (0.08) [0.96, 1.29] *p = 0.16*	0.034 (0.02) [-0.004, 0.07] *p = 0.08*
Suggestions	1.00 (0.02) [0.96, 1.04] *p = 0.99*	−0.008 (0.01) [-0.02, 0.004] *(p = 0.18)*	**1.043* (0.02) [1.00, 1.09]** *p = 0.046*	0.055 (0.03) [-0.001, 0.11] *p = 0.053*
Praise	0.984 (0.01) [0.96, 1.01] *p = 0.27*	−0.003 (0.01) [-0.02, 0.01] *p = 0.73*	0.974 (0.05) [0.89, 1.07] *p = 0.57*	−0.003 (0.05) [-0.09, 0.09] *p = 0.94*
# Provided				
Explanations	1.023 (0.01) [1.00, 1.05] *p = 0.07*	**0.019* (0.01) [0.003, 0.03]** *p = 0.02*	1.052 (0.04) [0.97, 1.14] *p = 0.20*	**0.07* (0.03) [0.01, 0.12]** *p* = 0.02
Suggestions	**1.018*** (0.003) [1.00, 1.02]** *p = 0.000*	0.007 (0.01) [-0.01, 0.02] *p = 0.26*	**1.11*** (0.03) [1.05, 1.18]** *p = 0.000*	0.04 (0.03) [-0.02, 0.09] *p = 0.17*
Praise	1.013 (0.01) [1.00, 1.03] *p = 0.07*	0.011 (0.01) [-0.002, 0.03] *p = 0.11*	0.932 (0.06) [0.83, 1.05] *p = 0.24*	−0.008 (0.03) [-0.07, 0.06] *p = 0.82*
# Revisions	—	**0.029*** (0.01) [0.02, 0.04]** *p = 0.000*	—	0.006 (0.01) [-0.02, 0.03] *p = 0.60*
Controls				
School title (Title I = 1)	0.817 (0.27) [0.43, 1.55] *p = 0.54*	0.15 (0.12) [-0.09, 0.39] *p = 0.22*	0.727 (0.25) [0.38, 1.41] *p = 0.35*	0.08 (0.07) [-0.07, 0.22] *p = 0.30*
Task 1 score	1.046 (0.07) [0.92, 1.19] *p = 0.49*	**0.38*** (0.09) [0.20, 0.57]** *p = 0.000*	0.994 (0.09) [0.83, 1.20] *p = 0.95*	**0.46*** (0.07) [0.32, 0.60]** *p = 0.000*
Gender (Female = 1)	1.049 (0.11) [0.86, 1.28] *p = 0.64*	−0.006 (0.06) [-0.12, 0.11] *p = 0.93*	1.06 (0.17) [0.78, 1.44] *p = 0.70*	−0.02 (0.08) [-0.17, 0.13] *p = 0.81*
AIC	1930.39	713.09	2026.62	743.75
BIC	1946.01	763.86	2042.24	794.51

Note: $N = 367$. *=$p < 0.05$, **=$p < 0.01$, ***=$p < 0.001$, — = not included in model.
1. Estimates for Model 1 and Model 3 were exponentiated and expressed as incidence rate ratios (IRR).
2. Robust standard errors (SE) are used to adjust for clustering at the teacher level.

4.2. Direct relationships of feedback activities to learning (RQ2)

The number of high-level revisions, but not low-level revisions, significantly predicted learning, partially confirming H2a (see Models 2 and 4 in Table 2). Number of provided explanations significantly predicted learning for both high-level and low-level aspects of writing (confirming H2b). H2c, that learning would be directly predicted by the number of provided suggestions, was not supported. The number of received comments was never a significant direct predictor of learning. Note that the 95% Confidence Intervals for the predictor estimates were larger for low-level revisions and learning than for high-level revisions and learning, suggesting that power may have played a role in producing fewer significant low-level effects. According to the sensitivity analyses, the results of the primary analyses concerning high-level learning were robust across models removing control variables and non-significant key predictors (see Table H2 in Appendix H). When it came to low-level learning, the regression coefficient estimate for number of provided explanations changed from marginal to significant in all the other models in the sensitivity analyses while still having a very similar effect size (see Table H4 in Appendix H), suggesting a power issue in the full model. Thus, number of provided explanations for low-level issues was considered as a meaningful predictor of low-level learning. All other estimates for low-level learning were stable in the sensitivity analyses.

4.3. Indirect relationships of feedback activities to learning via revision (RQ3)

The mediation analyses involved five control variables (Task 1 score, School title, Gender, # of received praise comments, # of provided praise comments), two independent variables of interest (# of received explanations, # of provided suggestions), one mediator (# of revisions), and one outcome (Task 2 scores). Only the significant predictors of number of revisions (i.e., number of received explanations, number of provided suggestions) were included. Similarly, mediation analyses were only conducted for the high-level aspects of writing because the number of low-level revisions did not significantly predict low-level learning. In these mediation analyses, both of the feedback variables were found to be indirect predictors of learning via revisions (see Table 3). The student-level indirect effect of received explanations on learning via high-level revisions was significant, indirect effect = 0.007; $SE = 0.0026$, 95% CI = [0.0024, 0.0124]; $p < 0.01$. The student-level indirect effect of provided suggestions on learning via high-level revisions was also significant, indirect effect = 0.003; $SE = 0.0013$, 95% CI = [0.0008, 0.0059]; $p < 0.05$. Thus, both number of received explanations and number of provided suggestions appeared to be beneficial for learning via their effects on revision, partially confirming H3b and H3c because the mediation pathways were not supported for low-level aspects of writing.

5. Discussion

Statistically significant relationships are summarized in Fig. 2, separately for high-level and low-level aspects of writing. To support reasoning about relative effect sizes, effect sizes of each relationship is presented on the figure in standardized units: the increase in incidence rate of revisions or the standard deviations increase in the transfer task for each standard deviation increase in the predictor. Overall, the pathways representing constructive learning opportunities had the strongest relationship to revision and learning outcomes, and the active pathway (received suggestions that were enacted in revisions) also had some support. However, the passive pathway (received feedback directly to learning) was not supported. Thus, this study adds to the previous research by showing differential effects of specific feedback activities on revisions and learning along ICAP lines.

Table 3
Within-direct and Within-indirect Effects of Received Explanations and Provided Suggestions on High-level Learning (i.e., Task 2 score controlling for Task 1).

	Within-direct effect (SE)	Within-indirect effect (SE)	Indirect effect (95 % CI)
Received explanations	−0.0189 (0.009)	**0.007 (0.0026)**	(0.0024, 0.0124)
Provided suggestions	0.0107 (0.0049)	**0.003 (0.0013)**	(0.0008, 0.0059)

Note: N = 367. Boldface type indicates a significant indirect effect / confidence interval not including 0.

5.1. Relationship of feedback activities to revisions (RQ1)

High-level aspects of writing.

It is not surprising that number of received explanations significantly predicted revisions (H1a, see Fig. 2A) because explanations have been shown to help clarify problems and persuade students to revise (Huisman et al., 2018; Strijbos et al., 2010; Wu & Schunn, 2020a). However, negative effects of received explanations on feedback implementation (e.g., Nelson & Schunn, 2009) and quality of revised drafts (e.g., Tseng & Tsai, 2007) have also been observed. The quality of received explanations may be an important moderator—the current study focused on students in an Advanced Placement course using a carefully designed rubric and peer review training before writing, which could produce a higher average explanation quality. The finding that the number of provided suggestions significantly predicted revisions (H1d) is also consistent with prior research that has demonstrated the more suggestions assessors give, the better they perform in their revised drafts (Lu & Law, 2012). In other words, the skill of providing suggestions to others could be transferred to and encourage assessors' own revisions.

The relatively small effect sizes for each of the statistically significant effects deserves some discussion. The regression approach used in this study has a number of methodological features that are likely to produce small effect size estimates: measurement noise from using only one writing task at pre and post; examining learning at the level of overall revisions and task scores, rather than within more specific areas of feedback and revision; and measuring learning from only one round of feedback and revision. Given these considerations, large effect sizes were unlikely and the observed coefficients were likely underestimates of the amount of learning that took place. However, it is also possible that additional, theoretically-important sources of variance may have played a role and those should be examined in future research. For example, the quality of explanations or suggestions was not measured, and low quality explanations and suggestions are likely a poor source of learning. In addition, this study focused entirely on the overt process of peer feedback, but covert processes might also have played a role: whether alternative revisions were considered by assessors in addition to the ones provided in the comment; whether authors not only read received comments but also evaluated them carefully enough to consider specific implications/next steps; and whether authors read comments just from a document revision perspective or also from a future task perspective. Qualitative data such as think-aloud protocols and interviews could be collected to elicit information on such covert processes.

Finding that the number of received suggestions were not predictive of the number of revisions (H1b) does not match earlier research. For example, Nelson and Schunn (2009) found that received suggestions were positively associated with implementation, while Tseng and Tsai (2007) found an in-between result: received suggestions were only useful for improving quality of revised drafts in the initial stage rather than later stage of the revising process. We noted that the current study focused on the overall number of revisions, rather than the more frequently-examined rate of implementation of actionable comments or quality of revised drafts. Overall revisions each student made included those that resulted from received and provided feedback as well as revisions triggered by other sources (e.g., proof-reading or automatic identification of spelling or grammar issues by Microsoft Word). Investigating received feedback features and implementation at the comment level explores a more direct relationship between received feedback features and implementation (e.g., received solutions and feedback

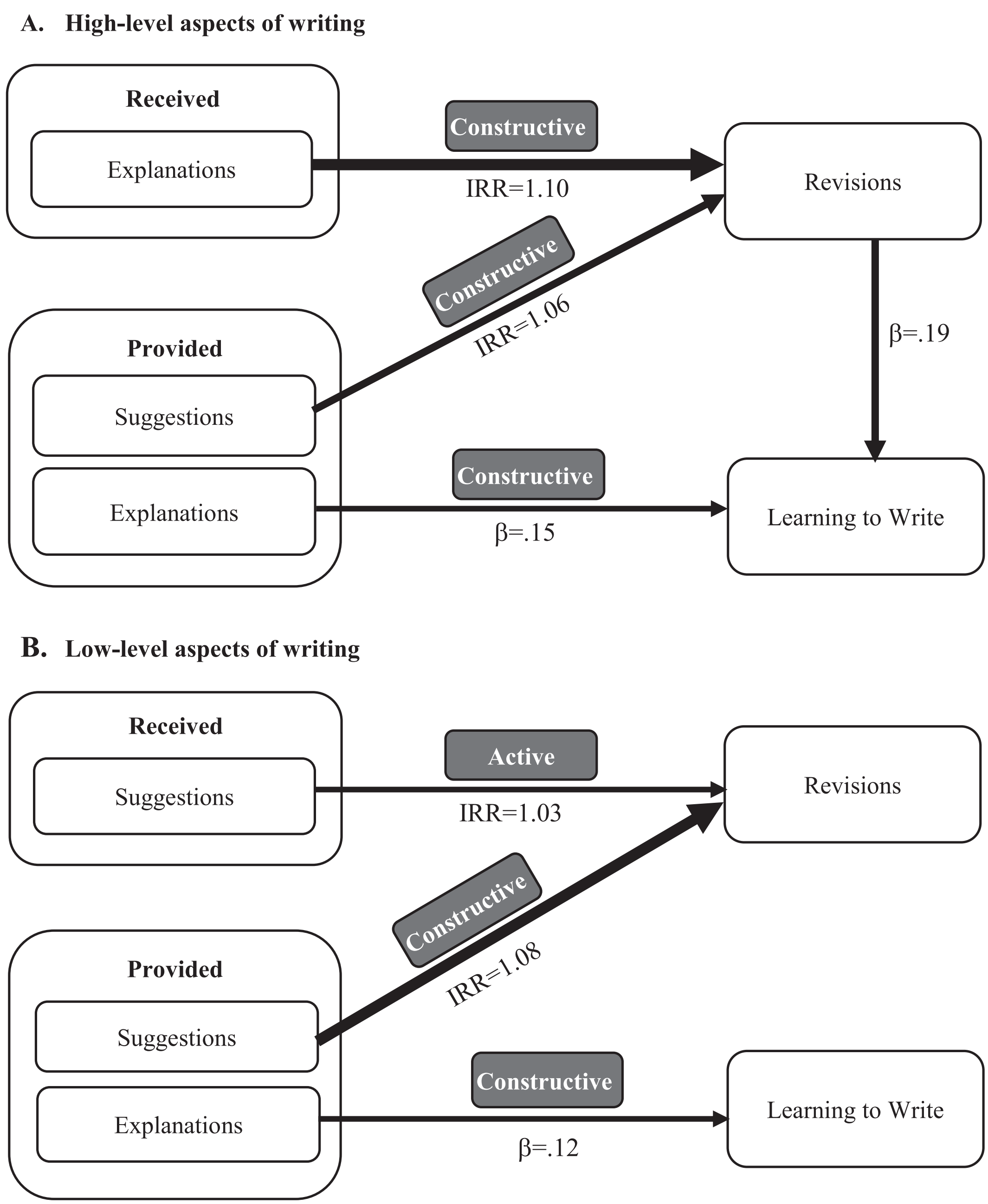

Fig. 2. Revised feedback-to-learning model for learning in high-level and low-level dimensions of writing, showing only statistically significant relationships, for analysis by feedback features. Numbers represent the incidence rate ratios per 1 SD change in the predictor from the negative binomial regressions and standardized coefficients from the linear regressions (see Table 2). Line thickness is proportional to the effect sizes of the relationships.

implementation), but ignores other revisions. Receiving more suggestions may lead to more passive revising overall, thereby failing to increase the overall number of revisions. The lack of a significant relationship between the number of received suggestions and total revisions may also be that students made overall revisions without using some suggestions as some suggestions might be incorrect, or not sufficiently specific. When the proposed suggestions are specific, students' attention can be focused and feedback can be more directed (Hattie & Timperley, 2007). Conversely, when a suggestion is not specific, students are more likely to ignore it, because they are more likely to complete tasks that require minimal effort (Luttrell et al., 2010).

Low-level aspects of writing

The number of received explanations did not significantly predict the number of low-level revisions (H1a) perhaps because only high-level problems are sufficiently complicated so as to need explanatory information. Lu and Law (2012) also found that explanations related to language use were not a significant predictor of revision. By contrast, the number of received suggestions was found to be a significant predictor of the number of low-level revisions (H1b, see Fig. 2B). This finding is consistent with prior research (Nelson & Schunn, 2009; Tseng & Tsai, 2007). Low-level problems are less cognitively complex than high-level problems, and therefore easier to solve with received suggestions, whereas students might rely on other sources than received suggestions for revising high-level issues. In addition, suggestions on low-level problems might be specific, so that the low-level revisions were more likely to be made with minimal effort (Luttrell et al., 2010). Like with high-level aspects of writing, provided suggestions did predict low-level revisions (H1d). It may be that the more constructive learning involved in figuring out how to solve problems in others' essays was useful to solving problems in the assessor's own essays.

5.2. Direct relationships of feedback activities to learning (RQ2)

High-level aspects of writing

The relationship of feedback activities to learning also varied across high-level versus low-level dimensions. As predicted, both number of revisions (H2a) and number of provided explanations (H2b) appeared to be directly related to high-level learning outcomes. The role of revisions in learning to write is well supported in the literature (Kellogg & Whiteford, 2009; Wu & Schunn, 2020b). Similarly, providing explanations is broadly supported as a learning mechanism in general (Aleven & Koedinger, 2002; Chi et al., 1994), as well as for learning to write in particular (Van Popta et al., 2017). While the studies of self-explanation are perhaps more widely cited and discussed, that research has acknowledged that "the advantages gained by explaining to others and to oneself are comparable" (Chi et al., 1994, p. 441). Mechanistically, Chi et al. (1994) argued that the more explanations students generate, the deeper their understanding because explaining enables students to construct declarative or procedural knowledge and integrate newly learned knowledge with prior knowledge. Further, when students explain their ideas explicitly, it provided opportunities for others to engage with the ideas by questioning, challenging, comparing, and connecting multiple ideas (Webb et al., 2021). We note that the observed relationship between provided explanations and the learning outcome was marginally significant and in need of replication in future research. The strength of this relationship might have been influenced by the quality of the provided explanations (e.g., poor explanations are likely a worse learning opportunity). It might also be influenced by ceiling effects on the outcome measures for those providing explanations: higher ability authors provided more explanations but had less room to improve on the outcome measure.

Hypothesis H2c (i.e., number of provided suggestions directly predicted learning) was not supported as a direct effect perhaps because provided suggestions could only be transferred to the assessor's own writing by routine practice (Wu & Schunn, 2020b). In other words, students learn to write by practicing their suggested strategies in their own revisions. Indeed, a significant indirect relationship between number of provided suggestions and learning via revisions was observed.

Low-level aspects of writing

For low-level aspects of writing, only the number of provided explanations were a significant predictor of learning (H2b). Why did number of revisions not predict low-level learning (H2a)? It is unlikely that students did not learn at all from revisions, especially given that the number of provided suggestions significantly predicted low-level revisions. Instead, it may be the relative gains from a single instance of practice were small—too small to be statistically detectable in the current study. In addition, there may an alignment problem given the wide range of possible specific low-level writing errors. Students are more likely to revise when the feedback they receive and provide align (Zhang et al., 2017). However, low-level revisions may be less likely to happen than high-level revisions because what students receive and provide on low-level issues are less likely to overlap. A longitudinal study examining the cumulative benefits of more writing tasks (and therefore more total revising and resulting learning) could be conducted to determine whether consistently making low-level revisions leads to low-level learning. Or perhaps there was no learning effect because low-level revisions could be done somewhat mindlessly.

Similarly, provided suggestions did not directly predict low-level learning (H2c). As noted above, low-level problems are likely more unique to each author whereas high-level problems are more likely to be shared across authors, so that students might learn more from providing high-level suggestions than from low-level suggestions (Wu & Schunn, 2020b).

5.3. Indirect relationships of feedback activities to learning via revision (RQ3)

The two significant predictors of high-level revisions, number of received explanations (H3b) and number of provided suggestions (H3c), were found to predict learning indirectly through revisions. It might be that students learn to write by figuring out how to solve problems after receiving explanations or practicing their own strategies in revisions, rather than directly learn from receiving explanations or providing suggestions. The positive effects of receiving explanations were observed in some prior research (e.g., Gielen et al., 2010; Huisman et al., 2018), but other studies have observed negative effects of receiving explanation on learning outcomes (e.g.,Tseng & Tsai, 2007). The quality of peer feedback might be an important factor explaining the different findings. Prior work examining whether the comments in the currently studied context tended to be accurate (i.e., leading to improvements in the document if followed) found that most of the peer feedback was accurate (see Wu & Schunn, 2020c), perhaps reflecting the well-structural peer review training, detailed and student-friendly rubrics, well-specified writing task, anonymous and efficient online multi-peer feedback. Other contexts with fewer supports for comment accuracy might result in less accurate feedback, which would then be less likely to lead to revision or improvements in later writing assignments.

That the number of received suggestions appeared to not be indirectly related to learning outcomes (H3a) is consistent with prior research. For example, Gielen et al. (2010) found that presence of suggestions did not improve student writing performance significantly across three different writing tasks. This pattern of findings is clearly predicted by the ICAP framework: making revisions after receiving suggestions is active but not necessarily constructive because students do not need to generate new knowledge to revise based on a received suggestion. By contrast, activities such as making revisions after receiving explanations or providing suggestions are constructive in that students are more likely to need to generate new knowledge when translating explanations to revisions or transferring suggested revisions to their own writing. Engaging in constructive learning activities promotes "deeper understanding that might transfer" (Raković et al., 2020, p. 2).

Making revisions plays an important role in learning to write. Making revisions after receiving explanations or providing suggestions allows students to make inferences of conceptual knowledge of writing, explain and apply procedural knowledge of writing, which makes their knowledge of writing more complete and correct, deepens their understanding of writing, and consequently facilitates learning transfer. Further, that the number of revisions predicted learning might be related to students' motivation, i.e., students who tend to make a lot of revisions will be more likely to produce better essays. Research is emerging that peer review activities can also influence students' writing motivation (Li et al., 2021; Wu & Schunn, 2021).

Different findings for high-level and low-level aspects of writing indicated the necessity of analyzing the two aspects separately. When feedback cues lead to elaborate or deeper processing of content, then learning should increase (Butler & Winne, 1995). However, compared with high-level feedback, low-level feedback (e.g., *In your second, third, and fifth paragraphs you start the paragraph with "The author". It gets repetitive just having "The author" start it off all the time, and is poor way to start a paragraph.*) may not require students to process knowledge as deeply as for high-level feedback. In general, high-level feedback tends to be more information-loaded, requires greater cognitive effort, and therefore is more likely to promote the internalization of correct knowledge. In other words, students' reflective process is likely not be activated to similar degrees for both high-level and low-level issues. However, it should be acknowledged the observed statistical power for the high-level analysis was larger than for the low-level analysis as indicated by the narrower confidence intervals on the effect estimates.

In line with the general ICAP framework predictions (Chi & Wylie, 2014) and prior studies using ICAP as the theoretical framework in other fields (e.g., Chase et al., 2019; Chi et al., 2018), the present study supports the relatively central role of constructive activities rather than passive and active activities in learning. When students engage in constructive learning, they connect new concepts to old concepts by monitoring their comprehension, elaborate on the nature of the problem, produce information that goes beyond what is presented to them explicitly, and deepen their knowledge (Chi et al., 1994; Nicol, 2012; Roscoe & Chi, 2008). The new knowledge learned from constructive learning can be used in new writing tasks (Nicol, 2014; Roscoe & Chi, 2008).

6. Conclusions

The main purpose of the study was to investigate the effects of different peer feedback activities on learning to write in a new writing task with a sufficiently large sample size to tease apart their individual effects. Although prior research has already revealed the effectiveness of peer feedback, the present study is the first to investigate the underlying mechanisms through which students learn from different types of peer feedback activities from the perspective of cognitive engagement and learning. We found support for both direct-only (provided explanations) and indirect-via-revisions-only (received explanations, provided suggestions) pathways, highlighting the importance of separately testing the underlying pathways. To help students engage in peer feedback more actively and productively, a teacher needs to implement peer feedback learning activities that stimulate learning. The present study provides teachers with specific suggestions on how to design and implement activities to optimize the benefits of peer feedback.

6.1. Implications for practice

The findings provided support for using multi-peer feedback in secondary school writing instruction. An important message for practice is that students should be encouraged to engage in the more constructive peer feedback activities such as providing explanations/suggestions and making revisions, especially based upon feedback with explanations rather than only feedback with specific solutions. Teachers should encourage students to make revisions of the same writing task, especially high-level revisions, rather than only assigning multiple writing tasks without any required revisions.

Providing good explanations can be challenging for students, and therefore it likely requires specific attention from teachers. Peer review training based on well-designed rubrics should include sessions teaching students how to explain problems clearly and various ways of providing suggestions. For example, students could be required to focus each feedback point on a single problem; questions could be included in rubrics, such as what the problem is, why it is a problem, and what examples can be used to support the explanation.

Task 1 score, an approximate measure of students' writing ability, is an important control variable that might influence how students perform and learn in peer review. Students with lower task 1 scores may require more support. Therefore, experimental research dividing students into different groups according to their task 1 scores should be conducted to uncover how task 1 scores influence the benefits students receive from participating in various components of peer review.

Further support for students could also be created through additional functions embedded into online peer assessment systems. For example, feedback features classifiers can be developed to remind assessors to include key feedback features when they fail to do so (Nguyen et al., 2017; Ramachandran et al., 2017). Alternatively, functions that enable assessors and assessees to talk about feedback and revisions interactively (but anonymously) could be developed, so that problems in received feedback (e.g., ambiguity, correctness) could be discussed and minimized.

The current study focused on evidence-based analytical writing and used genre-specific rubrics, which drew student attention to particular skills to practice. The findings may not generalize to other genres or situations using more generic peer feedback rubrics, where the peer feedback may be less focused on specific writing skills and instead focus on conceptual issues involved in the writing assignment topic. Future research should compare the impacts of using genre-specific rubrics vs general rubrics on peer review comments, revisions, and subsequent learning.

The research findings could also be different if a different peer review platform was used. The shared core features of most online peer review platforms help to make peer review effective and efficient, but different platforms contain various extra functions that shape student reviewing behaviors in different ways. For example, Calibrated Peer Review (CPR) has been found to help students learn in a variety of science courses (e.g., Hartberg et al., 2008) and writing (e.g., Russell et al., 2017). One of its significant features is calibration, which provides peer review training at the initial stage of peer review focused on the ratings that are generated. Instead of requiring a calibration phase, the online peer review system used in the present study, Peerceptiv, invites assessees to back-evaluate the helpfulness of peer feedback they receive, which shapes the grades assessors receive, and therefore improves the quality of peer feedback comments through this accountability mechanism. One study found that accountability that emphasizes the accuracy of ratings tends not to improve comment quality whereas accountability that emphasizes the helpfulness of comments does (Patchan et al., 2018).

6.2. Limitations and future research

There are several limitations to acknowledge in the current study. First, the current study was fundamentally correlational in nature, although with important controls included in the regressions. Further investigations using experiments should be conducted to compare the effects of the highlighted feedback activities on learning (e.g., active learning vs constructive learning, different paths of constructive learning such as learning by providing explanations vs learning by making revisions after receiving explanations). The current study provided the foundational correlation evidence to motivate key experimental details (e.g., which variables to emphasize and how they should be defined).

Second, interactive learning activities were not included in the study context given the (relatively common) form of peer assessment that was used. In written peer feedback, assessors and assessees do not normally interact with each other. However, according to the ICAP framework, students learn most deeply when they have activities that involve substantive interaction with one another. The interactive mode enables a student to collaborate with peer(s) in "co-constructing while dialoguing" (Chi et al., 2017, p. 12). Future research should directly test this assumption that interactive peer review activities do indeed produce even stronger learning outcomes than constructive peer review activities. For example, assessees could be allowed to have an online discussion with assessors, in which they could defend and challenge comments. Alternatively, students could collaboratively review a document rather than completing each of the multi-peer reviews alone (Mandala et al., 2017).

Third, students might have engaged in overt learning activities that were not captured in the electronic data sources. For example, students might have highlighted some of the feedback in printouts of the peer feedback they received (i.e., changing the activity from passive to active), or they might have discussed the feedback they received with peers (i.e., changing it from passive to interactive). Because the system interface does not make such actions easy to do, it is unlikely that it commonly occurred in the current study. But future work could specifically encourage and study the effects of such forms of active and interactive learning activities.

Fourth, only observable peer feedback activities were analyzed. However, students might have covertly engaged in active or constructive learning for the activities that did not require it overtly. For example, students could have reflected on potential revisions after receiving feedback without making revisions or explaining to themselves covertly. Students may also have developed other possible solutions in revisions after receiving suggestions. Future research could collect data on unobservable learning activities by means of surveys, stimulated-recall interviews or think-aloud protocols to directly test whether the focus on overt activities (as is typically done within ICAP) is a good approximation of internal learning processes in this context.

Fifth, quality of the peer feedback was not considered, by which we mean to what extent implementing the peer feedback will improve the paper. Feedback quality defined in this way not only influences how students respond after receiving feedback (Wu & Schunn, 2020c), but could also impact to what degree students learn from providing feedback. That is, students are more likely to make revisions after receiving high quality feedback, and they may be more likely to learn from providing high quality feedback. The current dataset likely had relatively few low-quality comments (i.e., not likely to produce document improvements if followed), based on comment quality coding applied to a partially overlapping dataset (Wu & Schunn, 2020c). Thus, it is unlikely that the core findings would change if quality had been included in the analysis. But, future research could explore whether the act of producing suggestions and explanations improves learning of the assessee and assessor because it requires deeper reflection and therefore results in higher quality comments.

Similarly, feedback style, a macro characterization of feedback (e.g., characterized as authoritative, interpretative, explorative, or collaborative, Leenknecht & Prins, 2018; van den Berg, 2003) could also be explored in future research using an ICAP lens. We note that these styles are often categorized using similar feedback features to our coding scheme (e.g., whether comments include positive/praise elements, suggestions for improvement, explanations). Therefore, as an initial hypothesis, we conjecture that the styles would mimic our current findings based upon the more micro-level feedback features they involve. For example, an authoritative style would have reduced learning for the assessee because it would produce passive or merely active learning and moderate learning for the assessor to the extent that it included suggestions. Interpretative and explorative styles might produce greater learning for both assessee and assessor if they generally tended to include explanations in addition to suggestions. Finally, a collaborative style (between assessee and assessor) might produce the greatest benefits for both because it would engage both in interactive learning. Alternatively, feedback style might have additional impacts via motivational effects on either assessor or assessee.

Finally, student engagement in revisions after receiving an explanation of a suggestion would vary according to the nature or complexity of the suggestion. Translating an explained complex suggestion in revisions demands more cognitive processing and are more challenging than using an explained simple suggestion. Measurement variability could be reduced in future research by further dividing explanations, such as dividing into explanations of more and less complex problems or explanations of problems vs explanations of suggestions.

6.3. Conclusion

In sum, students benefited from peer feedback in different ways for high-level versus low-level aspects of writing. For high-level aspects of writing, students appeared to learn to write via constructive forms of learning activities: making revisions using received explanatory information or provided suggestions, or by providing explanatory feedback to peers. For low-level aspects of writing, students tended to make more revisions on the basis of leveraging received suggestions and provided suggestions, or learn by providing explanations to peers. The findings of this study provide support for teachers to design and implement different constructive learning activities depending on which aspects of writing they are targeting. Teachers should encourage students to provide more explanations and make more revisions after receiving explanations and providing suggestive feedback. During training on peer review, students should be taught how to provide good explanations based on student-friendly rubrics.

Funding

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

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2023.102160.

References

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

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147–179. https://doi.org/10.1016/S0364-0213(02)00061-7

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

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

Bangert-Drowns, R., Kulik, C. L., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in testlike 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

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Chase, C. C., Marks, J., Malkiewich, L. J., & Connolly, H. (2019). How teacher task guidance during invention activities shapes students' cognitive engagement and transfer. *International Journal of STEM. Education, 6*(14). https://doi.org/10.1186/s40594-019-0170-7

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105.

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., et al. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science, 42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477. https://doi.org/10.1207/s15516709cog1803_3

Chi, M. T., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *The Journal of the Learning Sciences, 26*(1), 10–50.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*, 629–643. https://doi.org/10.1007/s11251-010-9146-1

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

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

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

DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretic foundations and new directions. *Review of Educational Research, 58*, 119–149. https://doi.org/10.3102/00346543058002119

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review, 32*, 481–509. https://doi.org/10.1007/s10648-019-09510-3

Fox, J. (1991). *Regression diagnostics: An introduction*. Newbury Park, CA: Sage.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304–315. https://doi.org/10.1016/j.learninstruc.2009.08.007
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. https://doi.org/10.3102/003465430298487
Hartberg, Y., Gunersel, A. B., Simspon, N. J., & Balester, V. (2008). Development of student writing in Biochemistry using Calibrated Peer Review. *Journal of the Scholarship of Teaching and Learning, 2*(1), 29–44.
Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanism. *American Behavioral Scientist, 64*(1), 19–54. https://doi.org/10.1177/0002764219859633
Hilbe, J. M. (2007). *Negative binomial regression*. Cambridge University Press.
Hu, G., & Lam, S. T. E. (2010). Issues of cultural appropriateness and pedagogical efficacy: Exploring peer review in a second language writing class. *Instructional Science, 38*, 371–374. https://doi.org/10.1007/s11251-008-9086-1
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 Higher Education, 36*, 955–968. https://doi.org/10.1080/02602938.2018.1424318
Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250–266. https://doi.org/10.1080/00461520903213600
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. https://doi.org/10.1037/0033-2909.119.2.254
Leenknecht, M. J. M., & Prins, F. J. (2018). Formative peer assessment in primary school: The effects of involving pupils in setting assessment criteria on their appraisal and feedback style. *Eur J Psychol Educ, 33*, 101–116. https://doi.org/10.1007/s10212.017.0340.2
Leijen, D. A. J. (2017). A novel approach to examine the impact of web-based peer review on the revisions of L2 writers. *Computers and Composition, 43*, 35–54. https://doi.org/10.1016/j.compcom.2016.11.005
Li, H., Bialo, J. A., Xiong, Y., Hunter, C. V., & Guo, X. (2021). The effect of peer assessment on non-cognitive outcomes: A meta-analysis. *Applied Measurement in Education, 34*(3), 179–203. https://doi.org/10.1080/08957347.2021.1933980
Liou, H. C., & Peng, Z. Y. (2009). Training effects on computer-mediated peer review. *System, 37*(3), 514–525. https://doi.org/10.1016/j.system.2009.01.005
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
Lu, J. Y., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*, 257–275. https://doi.org/10.1007/s11251-011-9177-2
Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*, 30–43. https://doi.org/10.1016/j.jslw.2008.06.002
Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2010). The mathematics value inventory for general education students: Developmental and initial validation. *Educational and Psychological Measurement, 70*, 142–160. https://doi.org/10.1177/0013164409344526
Mandala, M., Cole, E., Schunn, C., Goldberg, M., & Pearlman, J. (2017). Comparison of collective team and individual student peer feedback on design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers.
McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1
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.), *Feedback strategies for interactive learning tasks* (3rd ed., pp. 125–143). Mahwah, NJ: Erlbaum.
National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011* (NCES 2012-470). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
Nicol, D. (2012). Resituating feedback from the reactive to the proactive. In D. Boud, & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding it and doing it well* (pp. 34–99). London: Routledge.
Nicol, D. (2014). Guiding principles of peer review: Unlocking learners' evaluative skills. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 195–258). Edinburgh: Edinburgh University Press.
Nelson, & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401. https://doi.org/10.1007/s11251-008-9053-x
Nguyen, H., Xiong, W., & Litman, D. (2017). Iterative design and classroom evaluation of automated formative feedback for improving peer feedback localization. *International Journal of Artificial Intelligence in Education, 27*, 582–622. https://doi.org/10.1007/s40593-016-0136-6
Patchan, & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591–614. https://doi.org/10.1007/s11251-015-9353-x
Patchan, M. M., & Schunn, C. D. (2016). Understanding the effects of receiving peer feedback for text revision: Relations between author and reviewer ability. *Journal of Writing Research, 8*(2), 227–265. https://doi.org/10.17239/jowr-2016.08.02.03
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*, 2263–2278. https://doi.org/10.1080/03075079.2017.1320374
Patchan, M. M., Schunn, C. D., & Correnti, R. (2016). The nature of feedback: How feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103
Philippakos, Z. A., & MacArthur, C. A. (2016). The effects of giving feedback on the persuasive writing of fourth- and fifth-grade students. *Reading Research Quarterly, 51*(4), 419–433. https://doi.org/10.1002/rrq.149
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
Raković, M., Marzouk, Z., Liaqat, A., Winne, P. H., & Nesbit, J. C. (2020). Fine grained analysis of students' online discussion posts. *Computers & Education, 157*. https://doi.org/10.1016/j.compedu.2020.103982
Ramachandran, L., Gehringer, E. F., & Yadav, R. K. (2017). Automated assessment of the quality of peer reviews using natural language processing techniques. *International Journal of Artificial Intelligence in Education, 27*, 534–581. https://doi.org/10.1007/s40593-016-0132-x
Rittle-Johnson, B., Loehr, A., & Durkin, M. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM Mathematics Education, 49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z
Roscoe, R. D., & Chi, M. T. H. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science, 36*, 321–350. https://doi.org/10.1007/s11251-007-9034-5
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
Schunn, C. D. (2016). Writing to learn and learning to write through SWoRD. In S. A. Crossley, & D. S. McNamara (Eds.), *Adaptive Educational Technologies for Literacy Instruction*. NY: Taylor & Francis, Routledge.
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
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*, 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 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*, 1161–1174. https://doi.org/10.1016/j.compedu.2006.01.007
US Department of Education (DOE). (2018, October 24). *Improving basic programs operated by local educational agencies* (Title I, Part A). Retrieved from https://www2.ed.gov/programs/titleiparta/index.html.
Van den Berg, B. A. M. (2003). *Peer assessment in universitair onderwijs (unpublished doctoral dissertation)*. IVLOS: Utrecht University, Utrecht.
Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review, 20*, 24–34. https://doi.org/10.1016/j.edurev.2016.10.003
Webb, N. M., Franke, M. L., Johnson, N. C., Ing, M., & Zimmerman, J. (2021). Learning through explaining and engaging with others' mathematical ideas. *Mathematical Thinking and Learning*. https://doi.org/10.1080/10986065.2021.1990744
Wekerle, C., Daumiller, M., & Kollar, I. (2020). Using digital technology to promote higher education learning: The importance of different learning activities and their relations to learning outcomes. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2020.1799455
Wiggins, B. L., Eddy, S. L., Grunspan, D. Z., & Crowe, A. J. (2017). The ICAP active learning framework predicts the learning gains observed in intensely active classroom experiences. *AERA Open, 3*(2), 1–14. https://doi.org/10.1177/2332858417708567
Wooley, R. S., Was, C. A., Schunn, C. D., & Dalton, D. W. (2008). *The effects of feedback elaboration on the giver of feedback*. Washington DC: The Science Society.
Wu, Y., & Schunn, C. D. (2019). The learning science of multi-peer feedback for EFL students. *Technology Enhanced Foreign Language Education, 189*, 13–21.
Wu, Y., & Schunn, C. D. (2020b). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal, 58*, 492–526. https://doi.org/10.3102/0002831220945266
Wu, Y., & Schunn, C. D. (2020a). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology, 60*. https://doi.org/10.1016/j.cedpsych.2019.101826
Wu, Y., & Schunn, C. D. (2020c). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology, 62*. https://doi.org/10.1016/j.cedpsych.2020.101897
Wu, Y., & Schunn, C. D. (2021). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science, 49*, 365–394. https://doi.org/10.1007/s11251-021-09546-5
Wu, Y., & Schunn, C. D. (2022). Assessor writing performance on peer feedback: Exploring the relation between assessor writing performance, problem identification

accuracy, and helpfulness of peer feedback. *Journal of Educational Psychology, 115* (1), 118–142. https://doi.org/10.1037/edu0000768

Yu, S. (2021). Giving genre-based peer feedback in academic writing: Sources of knowledge and skills, difficulties and challenges. *Assessment & Evaluation in Higher Education, 46*(1), 36–53. https://doi.org/10.1080/02602938.2020.1742872

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