Instructional Science (2026) 54:59
https://doi.org/10.1007/s11251-026-09813-3

ORIGINAL RESEARCH

Beyond small gains: how content-focused high-specificity peer feedback correlates with substantial provider benefits

Jing Yang[1] · **Christian D. Schunn**[2]

Received: 30 August 2025 / Accepted: 30 June 2026

Abstract
Prior meta-analyses and meta-regression studies on peer feedback have consistently observed benefits for providers, but alongside notable and unexplained variability in effect sizes. While some students exhibit minor learning gains from providing feedback, others display significant improvements. We hypothesized that the features of students' comments may play pivotal roles in the variability of learning by providing. To test these hypotheses, we strategically selected 12 assignments from a comprehensive dataset generated in an online peer feedback platform: six assignments demonstrating weak learning outcomes and six strong ones. The study particularly focused on variation in near-transfer learning rather than improved performance in revised version of the same assignment. Peer comments were strategically selected controlling for comment *length* and coded on two key dimensions: *feedback focus* and *feedback specificity*. Our results revealed that assignments with weak learning outcomes displayed a predominance of lengthy form-focused comments. Conversely, assignments with strong effects were characterized by a prevalence of questions for clarification, suggestions for improvement, and direct revisions concerning content issues. Another notable finding was although the strong cases addressed more content issues than the weak cases did, the difference was not statistically significant, which indicated that the key factor was not simply the greater number of content-focused comments but their specificity. Emphasizing an integrated perspective, we proposed that the ratio of specific content-focused comments to form-focused comments was crucial for identifying situations conducive to student learning through providing feedback. Conceptually, these findings are explained by variation in behavioral engagement and cognitive engagement of the feedback provider.

Keywords Peer feedback · Learning gains · Provider · Engagement

✉ Jing Yang
yangjing@dlufl.edu.cn

[1] Department of Foreign Language Education for Non-Majors, Dalian University of Foreign Languages, 6 Lvshun South Road, Dalian 116044, China

[2] Learning Research and Development Center, University of Pittsburgh, 3420 Forbes Ave, Pittsburgh, PA 15260, USA

Published online: 06 July 2026

Introduction

Peer feedback is increasingly used as a pragmatic, valid, and effective pedagogical strategy for undergraduate students within a wide variety of courses and contexts (Yu & Schunn, 2023). Although peer feedback is consistently found to be an important learning opportunity, especially for feedback providers (Li et al., 2010; Lundstrom & Baker, 2009; Van Popta et al., 2017; Zhang et al., 2023), the learning benefits show very large variation across studies (Van Popta et al., 2017), courses (Zong et al., 2021), assignments (Yu & Schunn, 2023) and feedback types (e.g., cognitive comments vs. affective comments in Cheng et al., 2015; Lu et al., 2023).

Multiple meta-analyses have established that peer feedback benefits student learning in higher education (e.g., Huisman et al., 2019; Li et al., 2020; Yan et al., 2022). Peer feedback not only improves students' writing performance compared to no feedback or self-assessment (Huisman et al., 2018), but in some cases even yields larger effects than teacher assessment (Li et al., 2020). More recent meta-analyses have found that technology-facilitated peer assessment significantly enhances students' learning outcomes and positively impacts higher-order critical thinking (Zhan et al., 2023; Zheng et al., 2020).

Many studies have found that the act of providing feedback can be a powerful driver of learning, often more so than receiving feedback—but its benefits depend on how students engage in the process (Ion et al., 2019; Li et al., 2010; Patchan & Schunn, 2015). Li et al. (2010) provided direct empirical evidence that giving feedback improved students' own performance, even after accounting for their initial ability. Notably, receiving feedback did not show the same effect. Patchan and Schunn (2015) explained why providing feedback was beneficial by focusing on the cognitive processes involved. It suggested that giving feedback allowed students to practice essential skills such as identifying problems and proposing solutions. However, the benefits were not uniform—they depended on the reviewer's ability and the nature of the feedback produced. More skilled reviewers generated more critical, problem-focused, and solution-oriented comments, whereas less skilled reviewers tended to provide more superficial praise, limiting these benefits. Ion et al. (2019) showed that students themselves recognized the value of providing feedback. They reported improved learning experiences, greater engagement, and stronger commitment to both their own and their peers' learning.

Literature review

Learning variation

Despite broad agreement on the benefits of providing peer feedback, meta-analyses also report substantial variation in effect sizes and attempt to identify moderating factors. Larger effects have been observed when students combine quantitative ratings with qualitative comments (Huisman et al., 2019; Li et al., 2021), receive rater training (Li et al., 2020; Sanchez et al., 2017), use computer-mediated systems (Li et al., 2020; Yan et al., 2022), or provide feedback anonymously (Zheng et al., 2020).

However, these moderators explain only a small proportion of the variance across studies. For instance, Li et al. (2020) found that computer-mediated peer assessment explained

less than 3% of the variance in effect sizes. Similarly, Yan et al. (2022) tested several potential moderators but found significant effects only for online technology use, leaving most variation still unexplained. Regarding diverse age groups, other meta-analyses (e.g., Sanchez et al., 2017; Yan et al., 2022) reported that age and educational level did not significantly moderate outcomes either. Nevertheless, older students tended to show slightly larger (albeit non-significant) gains—likely due to greater metacognitive skills and ability to manage cognitive load needed for giving peer feedback.

Variation in learning benefits appears to be driven by other contextual moderators such as assignment characteristics, courses and disciplines, and institution types. For example, Wu and Schunn (2021) examined a shared sequence of assignments across multiple high schools and found variation across schools. Zong et al. (2021), analyzing 13 university courses, observed significant variation in effect sizes across courses. Yu and Schunn (2023), in a large-scale study of 243 courses, also identified substantial variation for feedback providers. Notably, the third study (Yu & Schunn, 2023) suggested that true variation at the course level was modest, with most differences occurring across assignments. In particular, within-course variation across assignments was greater than between-course variation. Further, Karim and Nassaji (2020) found that learning gains tended to be smaller when assignments required far transfer (i.e., applying knowledge to dissimilar tasks).

Feedback features

Although research on learning gains for feedback providers is relatively limited, existing studies have emphasized the importance of feedback features—such as problem identification, suggestion, explanation, and solution. These features appear to play a more critical role than contextual factors in driving writing improvement, particularly for feedback providers. For example, Li et al. (2010) found that students' ability to identify key issues and provide constructive suggestions predicted the quality of their final projects, whereas the feedback they received did not show a similar relationship. Cho and MacArthur (2011) also demonstrated that students who generated feedback outperformed those who only receive peer comments, particularly when their feedback included problem identification and solutions. Overall, features such as problem identification and explanation were consistently associated with higher implementation rates in revisions (Nelson & Schunn, 2009; Patchan et al., 2016).

Feedback features can target either content (e.g., ideas, argumentation, structure) or form (e.g., grammar, spelling, formatting). Although the content-form distinction has been critiqued (Hyland & Hyland, 2006), it remains useful. In general, when peer comments are featured with complex content-focused elements, receivers are more likely to make macro revisions (Cho & MacArthur, 2011; Patchan et al., 2016; Tuzi, 2004; Wu & Schunn, 2023). Even though receivers might choose to ignore content-focused feedback, when feedback providers are required to compare ideas, question arguments, suggest possible modifications to content elements, the providers are challenged into a meaning-making and knowledge-building process that is key to learning (Van Popta et al., 2017; Zhang et al., 2023). By contrast, when students provide extensive feedback on superficial form-focused elements such as grammar and word choice, the potential benefits of these feedback features may not become evident for providers themselves. Lengthy explanations of a grammar rule may not be equally beneficial as lengthy questions about ideas and arguments from a cognitive

perspective. To date, few studies have explored how these features (problem identification, suggestion, explanation, and solution) interact with feedback focus (content vs. form) in ways that either support or hinder the writing development of peer feedback providers.

In addition to focus, the specificity of feedback is another critical factor. Explanations and suggestions can vary from general to specific, and this distinction has important implications. For example, Wu and Schunn (2021) distinguished between general suggestions and specific solutions, showing differences in their impact on how recipients decide whether to implement feedback. Similarly, Huisman et al. (2018) found that specific explanations led to deeper processing and more meaningful revisions, whereas Ng et al. (2015) found vague suggestions were less effective. From the provider's perspective, deeper learning is more likely to occur when students engage in cognitively demanding processes by incorporating specific forward-oriented feedback into their feedback and giving step-by-step directions (Van Popta et al., 2017; Zhang et al., 2023). In contrast, general or vague comments are less likely to elicit such processing.

Cognitive and behavioral engagement

Variation in learning benefits from peer feedback has important theoretical implications. Several mechanisms have been proposed, including the development of domain knowledge (Huisman et al., 2018), increased motivation and agency (Ion et al., 2019), and enhanced metacognitive skills through critical engagement (Van Popta et al., 2017).

Research on cognitive engagement suggests that students who engage in constructive activities—those that require deeper cognitive processing—tend to learn more effectively than those who engage in purely behavioral tasks (Chi et al., 2018; Chi & Wylie, 2014). This aligns with Wu and Schunn (2023), who also argue that differences in cognitive engagement may explain variation in peer feedback outcomes.

In the present study we control for behavioral engagement by selecting cases with similarly long comments. The reason is that prior studies (Zong et al., 2021; Wu & Schunn, 2021; Yu & Schunn, 2023) found that longer comments were associated with greater learning gains. While carefully controlling many other experiences (e.g., amount of feedback received; prior performance differences), these studies consistently found significant positive relationships between comment lengths and learning benefits. Although comment numbers and lengths do not fully capture cognitive processes, they provide observable indicators of behavioral engagement.

The main focus of the present study falls on feedback providers' cognitive engagement, operationalized through the focus and the specificity of their comments. Affective engagement, while relevant for attitudes, generally explained less variance in learning outcomes than cognitive and behavioral engagement (Chen et al., 2023; Fan & Xu, 2020; Lu et al., 2023). Therefore, affective engagement is not within the scope of this study.

The present study

Research on peer feedback has highlighted the importance of providing feedback (rather than merely receive it) for undergraduates and suggested that feedback providers experience varying levels of learning benefits. In comparison with contextual factors, the differences

across assignments and the quality of feedback seem to matter more in driving writing improvement and meaningful learning. Content-focused, specific feedback tends to be more effective than form-focused, superficial feedback. Longer comments tend to be indicative of greater gains in peer feedback, too.

In principle, results of those studies seem pretty obvious: students learn more when they provide high quality feedback than when they provide superficial feedback, and this is because the former implies a greater cognitive engagement than the latter. However, this claim has not been rigorously tested in the context of transfer. Notably, most prior research of undergraduates' peer feedback has focused on revision tasks rather than new writing tasks. Therefore, it remains unclear whether such feedback facilitates performance on subsequent similar but novel tasks (i.e., near transfer).

Moreover, existing research has primarily examined the presence of learning gains, rather than the extent of their variation. When peer review yields differential learning benefits for providers, systematic comparisons between the feedback produced by those with greater-versus-smaller gains can illuminate the nuanced distinctions between these groups.

This study explores how feedback comments in assignments with strong learning effects differ from those in assignments with weak learning effects among undergraduate students. The study addresses the gaps by:

1. Focusing on variation in near-transfer learning rather than improved performance in revised version of the same assignment.
2. Comparing degrees of providers' learning gains (greater vs. smaller), rather than binary outcomes (with gains vs. without gains).
3. Examining how feedback features, focus, and specificity—while controlling for feedback length—interact, rather than treating them independently.

Accordingly, we hypothesize that:

1. Assignments with similarly long peer comments but a greater proportion of content-focused feedback will yield stronger learning benefits for feedback providers.
2. Assignments with similarly long peer comments but higher specificity will yield stronger provider learning benefits.
3. High-specificity, content-focused feedback will produce the strongest learning benefits compared to other combinations.

Methods

Assignment selection

The source dataset involved thousands of courses from hundreds of institutions worldwide using the Peerceptiv online peer feedback system. The dataset selected for the present study comprised 12 Peerceptiv assignments in 12 different courses from eight U.S. universities. In each given course, students as individuals completed multiple assignments and provided peer reviews to multiple peers. Students gave both ratings (1–7 on a number of instructor-created rubrics) and comments (according to instructor-created prompts). In each assign-

ment, we calculated each student's task performance based on the average score received from their peers; prior research has established strong reliability and validity of mean peer ratings in this larger dataset (Authors, 2023), mirroring strong general validity in peer review (Li et al., 2016).

How did these 12 assignments were systematically selected from the source dataset? How did they show either an especially strong relationship (six cases) or an especially weak relationship (six cases) between providing comments and learning? First, we used assignment-specific z-scores to account for differences across rubrics between courses and assignments. Specifically, we used a regression approach in which each student's z-score in the next assignment was predicted by the total number of words they provided in the current assignment. We controlled for their task performance z-score in the current assignment and two other salient controls (comment numbers and the perceived helpfulness of peer comments). Then, the standardized beta for the total length of provided comments was taken as the measure of learning from providing: a large standardized beta (i.e., 0.23 to 0.35) was taken as "strong", and a small standardized beta (i.e., 0.07 to 0.10) was taken as "weak" (see Appendix A for examples of strong and weak relationships).

Following the approach developed and validated (Wu & Schunn, 2021; Zong et al., 2021; Yu & Schunn, 2023), learning was conceptualized as the change in relative performance across assignments (e.g., moving from relatively weak performance in one assignment to relatively strong performance in the next). We only considered cases of *moderate transfer* between consecutive assignments (e.g., two different tasks involving the same genre, like two lab reports, two related genres, or a Methods section followed by a Results section); we ruled out cases of a first draft followed by a second draft (which involves changing task performance but not necessarily learning) or completely unrelated genre (which showed a weak relationship in performance across assignments and likely little evidence of improvements due to feedback providing).

Table 1 presents detailed descriptions of each selected assignment, ordered by discipline. In terms of assignment sequential number, the weak cases were more likely to involve the first-second assignments in a course, but at least half of both sets were second-third assignments. Further, the standardized beta generally showed no regular relationship to assignment number in the larger regression (Yu & Schunn, 2023). The strong and weak cases were also similar in terms of how many documents each student reviewed, how many dimensions had to be addressed in each review, and the overall number of words produced in each review (in the case of long comments). Appendix B presents more details on the selected cases, and Appendix C presents example rating rubrics and comment prompts for one selected assignment and its next assignment within that course.

The strong and weak cases were well matched on other dimensions, too. First, only university courses were considered as they were the majority case in the dataset (91%), and high school peer feedback tended to be much shorter. Second, consecutive assignments were individually submitted (i.e., not group assignments). Third, to improve generalizability, a similar range of disciplines was represented in both sets and a similar range of assignment genres were represented, including argumentative essays, lab reports, and summary writing etc. No course was repeated (one same course, if using Peerceptiv for different semesters, was chosen only once). Finally, the majority of the selected cases had at least 50 students, so as to have sufficient power in estimating providing effects at the level of a single

Table 1 For the selected assignments, the course discipline, the standardized beta for the relationship of provided comments and performance growth, the sequential assignment number within the course, the number of students in the regression, the number of documents each student reviewed, the number of reviewing dimensions, and the mean number of words within the selected comments

Case	Discipline	Provided beta	Assign. #	# of stud.	# of docs	# of dims	Mean # of words
Strong providing effect cases							
1	German	0.35	3	25	2–4	4	81
2	Anthropology	0.29	2	52	3–4	4	68
3	Biology	0.31	2	211	3–4	1	89
4	Biology	0.23	3	293	3–6	1	75
5	Biology	0.32	2	32	3–6	4	76
6	Engineering	0.25	3	49	1–5	6	95
Weak providing effect cases							
7	Communication	0.10	1	98	3	3	85
8	Biology	0.07	2	93	1–2	3	92
9	Biology	0.10	1	181	4–5	1	81
10	Biology	0.07	2	60	3–4	1	100
11	Engineering	0.10	2	56	3–4	7	63
12	Engineering	0.10	1	65	4–5	5	68

assignment. Some smaller courses (e.g., 25 students studying the German language) were included to improve the aforementioned generalizability.

Comment selection

The number of provided comments in each assignment varied greatly (from 379 to 2,268 comments per assignment) for a variety of reasons, but especially because of the large variation in class size. It was not feasible to exhaustively code all of the comments, nor was it necessary, based upon the law of large numbers. Therefore, we started focusing upon the longer comments in each assignment, since these were the comments that particularly drove learning, and we hypothesized the nature of the long comments (rather than the short comments) would be the critical varying element between the strong and weak providing effect cases.

How a random of 50 comments were selected from each assignment? First, An MSExcel formula was used to count the words in each comment and identify comments with 50 or more words. This threshold for defining long comments was based upon prior research which looked at the relationship between comment helpfulness and comment length (Zong et al., 2021). Second, using this threshold, we found each selected assignment typically had several hundred long comments that could be coded. The range varied from 51 to 315 long comments in an assignment. Our next step was to select a random 50 comments using the rand function in MSExcel. This approach would produce an efficient estimate of long comments that was similarly precise in each assignment with roughly 5% standard errors. We hypothesized that small differences in comment characteristics (i.e., smaller than 10%) would be unlikely to cause differences in learning effects, and thus 50 comments would be sufficient to detect meaningful differences.

Notably, the method for comment selection made random decisions at the comment level and thus sometimes selected slightly more or slightly fewer than 50 cases. The final dataset

included 301 comments for the strong group and 298 comments for the weak group. Mean comment length among the selected comments was between 68 and 95 words for the six strong cases, and 63 to 100 for the weak ones.

Participants

The 12 assignments were NOT written by 12 students. To clarify, these assignments actually involved a total of over 1215 undergraduate students who each participated in only one of the selected assignments. Among them, only those whose comments met our aforementioned assignment-comment selection criteria were studied. Specifically, 373 students were finally included as authors for the sampled 599 comments for coding.

Measures

The comment coding unit generally involved a single sentence for the majority of the comments, but could also extend to a few sentences. Each unit was supposed to capture and cover one meaningful feedback point, including words providing context to the comment, but excluding words addressing a different topic. Each feedback point (unit) was then coded in terms of *focus* and *specificity* (see Table 2).

Comment focus. The three focus codes were *Content* (i.e., focused on meaning issues), *Form* (i.e., focused on communication conventions, such as English spelling and grammar issues), and *Others* (i.e., did not specify meaning or form strengths or weaknesses). For example, comments such as "I think you did a great job" was coded as *Others* because there

Table 2 Comment codes related to content and form comments and their descriptions

Focus	Specificity	Description
Content		Comments focused upon: – clarity/ambiguity of claim, statement, and argument – (in)adequacy of support evidence – paragraphs (or ideas in one paragraph) in a logical order – transition between ideas
	Not specific	No additional detail.
	Specific	The reviewer: – refers/responds to a specific point in the original text, – explains *in brief* the comment, *without* providing suggestions or direct revisions.
	Strongly specific	Following a comment on content/organization, the reviewer: – asks *multiple* questions to seek for clarification, – provides *lengthy and detailed* explanations (at least two meaningful points) of a focal issue, – or *revises* or *rewrites* directly for the writer.
Form		Comments focused upon: spelling, capitalization, punctuation, grammar, lexical accuracy and variety, syntactic accuracy and variety, citation format, word limits
	Not specific	No additional detail.
	Location	The reviewer points out a particular section/paragraph/sentence or directly quotes from the written text to indicate location.
	Identification	The reviewer indicates the types of errors.
	Solution	The reviewer directly provides corrections.
Others	Not specific	– Comments with no reference to the written text – Praise or criticism on effort

was no reference to whether the strength was in meaning or in form. All three focus codes could involve either corrective/weaknesses or non-corrective/strengths comments.

Comment specificity. There were various ways in which content and form comments could be specific. If content comments did not include any detail, they were coded as *Not specific* (e.g., "I like your ideas but more details could be added." "Something is wrong with your overall structure."). Content comments could also be *Specific*, referring to a particular line/paragraph in the original text or giving brief explanation to an issue. Finally, content comments were carefully coded as *Strongly Specific* when they matched three important features (see Table 2 for details). Likely, form comments were coded in terms of location, identification and suggested correction.

Below are two examples of coded comments. The first example has two different *Content* idea units, one unit coded as *Specific* and the other unit coded as *Strongly specific*. The second example has only one *Form* unit, which includes both *Identification* and *Location* information.

- *The impact statement was great [Content], because you did not simply mention agriculture, but the evolutionary impact as well [Specific]. Your methods lacked a lot of details [Content]. I would suggest including what exactly was going on: researchers that kept tabs on the types of bedbugs found in the U.S. were sent a sample that was obviously not the same species as what they knew was located in the states [Strongly specific].*
- *Grammatically [Form], a lot of your sentences were run-ons [Identification] and never really carried out a full thought. For example, in your hypothesis statement, you accidentally included "and" twice in a row [Location].*

As a result, each of the coded comments produced a count: the number of units that were *Not specific*, *Specific* or *Strongly specific* in the case of *Content* comment units, and the number of *Not specific*, *Explanation*, *Location*, or *Solution* in the case of *Form* comment units, as well as *Others*.

A second coder was used to establish inter-rater reliability (see Table 3). The *Identification* and *Solution* codes had lower reliability levels because they were relatively rare (and thus one error greatly affected reported reliability). Given that both coders agreed these were rarely occurring and the data analysis focused on relative frequency within an assignment, the reliability of those codes was deemed sufficient. Appendix D presents additional examples as well as the more complete coding manual.

Data analysis

Due to the structure of providers' comments nested within 12 different assignments/courses, Hierarchical Linear Modeling (HLM) was implemented using the mixed command in Stata 18. The purpose was to ensure that the effect was robust across the sampled courses. Since the courses and the comments were both sampled from a broad pool, it was unlikely that there were direct interactive effects of one student's comment on another student in that

Table 3 Inter-coding reliability Kappa values

	Content			Form				Others
	Not specific	Specific	Strongly specific	Not specific	Location	Identification	Solution	
$n=50$	0.79	0.81	0.89	0.79	1.00	0.38	0.59	0.75

class. However, without accounting for the nested structure of comments within courses, it is possible that a large effect in just a few courses would produce a statistically significant result overall, even though unique factors of those few courses was the actual cause.

In addition to the quantitative analysis, comments were selected from four strategically sampled students for an in-depth qualitative analysis of their feedback features. We initially focused upon six similar cases (all from biology, see Table 4). We then identified the top five reviewers in each case in terms of the total comment words, which resulted in 30 reviewers. Next, we examined all the comments each of these 30 reviewers produced. Finally, we presented comments from four reviewers from four courses (two strong, two weak) that were typical of the patterns reflected in the quantitative analysis results.

Results

Content comment differences

As predicted in Hypothesis 1, we found that the strong providing effect cases trended towards having more *Content* comments, but the difference was moderate in size and not statistically significant (see Table 4). In other words, it was not a simple matter of pointing out more content issues. Instead, there were clear differences in terms of comment specificity, as predicted by Hypotheses 2 and 3. The largest differences (and both statistically significant) were for *Specific* and the sum of *Specific* and *Strongly Specific* comments, with

Table 4 Mean occurrence of comment features within the randomly selected long comments within each assignment, along with means per group of cases, Cohen's *d* for the difference between strong and weak groups and the *p*-value of the difference

Case No.	Discipline	Content	Spec.	S.Spec.	Sum	Form	Loca.	Expl.	Sug.	Sum	Others
Strong providing effect cases											
1	German	1.35	0.92	0.53	1.45	0.22	0.24	0.14	0.08	0.45	0.16
2	Anthropology	1.19	0.85	0.17	***1.02***	***0.47***	0.43	0.32	0.26	***1.00***	0.28
3	Biology	2.00	***0.76***	0.29	***1.04***	0.20	0.08	0.18	0.02	0.29	0.53
4	Biology	2.45	1.45	0.33	1.78	0.27	0.06	0.04	0.04	0.14	0.88
5	Biology	2.22	1.22	0.65	1.86	0.04	0.00	0.02	0.00	0.02	0.18
6	Engineering	3.24	1.65	0.48	2.13	0.35	0.11	0.13	0.13	0.37	0.44
	Mean	*2.10*	*1.15*	*0.41*	*1.56*	*0.26*	*0.15*	*0.14*	*0.09*	*0.37*	*0.41*
Weak providing effect cases											
7	Communication	1.65	1.06	0.13	1.19	0.42	0.19	0.23	0.10	0.52	0.08
8	Biology	1.47	0.69	0.31	1.00	0.69	0.55	0.43	0.37	1.35	0.14
9	Biology	2.69	0.89	0.29	1.17	0.40	0.14	0.10	0.10	0.33	0.96
10	Biology	2.39	1.08	0.49	***1.57***	0.96	0.53	0.39	0.39	1.31	0.31
11	Engineer	1.58	0.72	0.28	1.00	***0.22***	0.08	0.12	0.06	0.26	0.12
12	Engineer	0.60	0.48	0.04	0.52	0.70	0.72	0.46	0.32	1.50	0.06
	Mean	*1.73*	*0.82*	*0.26*	*1.07*	*0.56*	*0.37*	*0.29*	*0.22*	*0.87*	*0.29*
	Cohen's *d*	0.5	**1.1**	0.9	**1.2**	**-1.5**	-1.0	**-1.1**	**-1.1**	**-1.1**	0.4
	p-value	0.39	**0.04**	0.08	**.03**	**0.006**	0.06	**.04**	**.04**	**.04**	0.16

BoldItalic font indicates cases that violate the group pattern on dimensions with significant differences

the strong cases tending to have almost 50% more comments that were either *Specific* or *Strongly Specific*.

To illustrate the focus on strongly specific content, we present typical comments observed in the qualitative case study. Below are two comments provided by one strong-effect reviewer from Case 4. These comments were not only highly content-focused but also strongly specific. For instance, there was "an additional point" in all the six comments the reviewer provided in that assignment. Key phases and words were in bold for emphasis.

- *I would first like to say that your impact statement was very thoughtful and insightful. It also provided additional information to consider why this research may be useful to human society. Preserving human food supply is very important! I appreciate that you went the extra mile and did additional research to back up your claim, which evidently made it stronger!* ***An additional point to be considered*** *is also that these beetles, although not present in North America, can be accidentally carried over here as well. This means that by using this research to potentially eliminate or lessen these beetles, we will also be able to potentially less the possibility of them ruining the crops here in the U.S Overall, you did an awesome job and you were very thorough!*
- *I thought that your impact statement was very insightful and provided vital information about how it directly impacts human society. I also like how you mention specific examples of how this could be detrimental to the ecosystem.* ***An additional point to be considered*** *is maybe how this beetle also affects agriculture, and how might controlling their mating could potentially help agriculture as well (if it does have an affect on that as well). Overall, I felt that you did a great job on this assignment!*

Another typical reviewer from Case 5 was also characterized by a focus on content (such as on explanation of research results and their connections to hypothesis). In fact, the total of 16 comments this reviewer provided to four different authors all included specific suggestions or explanations about content. Two comments are provided below.

- *I think it would have been effective if you referred to your graphs in this section. It would have been easier to understand their purpose. I think your explanations were well developed and the flow was nice. I would have liked have seen you connect the analyses to your hypothesis, and explain why the 65% Hct level was the most effective. I know you touch upon this in your conclusion, but this results section would have been a good place for you to do so.*
- *Very nice job in your introduction, all information presented is accurate, and it's a perfect introduction. I think it was very effective to separate your hypothesis into another section. This allowed you to address the purpose of the experiment while also stating your proposed hypothesis. You also went above and beyond by testing all three methods instead of just 2 (required). Very nice job.*

Form comment differences

Overall, as predicted by Hypothesis 1, the weak providing effect cases tended to have twice as many *Form* focused comments than did the strong providing effect cases (see Table 4). This also involved significantly more identifications of error types and suggestions for how

to correct those form errors, or combinations of various form-involved comment features. However, it is worth noting that one of the strong cases (Case 2) had relatively many form-focused features, and one weak case (Case 11) had relatively few form-focused comments and features. As predicted by Hypothesis 3, the dependency on specificity was not especially strong for form-focused comments.

To illustrate the focus on form in the weak providing effect cases, we present typical comments observed in the qualitative case study. For example, four out of nine comments from a reviewer in Case 8 were form-focused; the form issues were located, identified and given solutions in great detail. The other five comments, though content-focused, were not specific; "evidence" in one comment was mentioned six times, but had no specific references or explanations.

- *You did a good job organizing your paper. It flowed which was good. You had a few passive voices. And also a sentence that is really wordy. Ex: "This means that there is less than 1% difference that separate who is a human and who is a chimp." You can take some words out of that sentence and it will be more clear. Also, you should not start a sentence with but so instead i recommend you use words like however, nevertheless etc. Your thesis can use some work its pretty broad and just there try and make it engaging and specific. Also on this sentence: "The story starts with my friend and me discovering a frog in Putah Creek and picking it up, examining it carefully." I think it is my friend and I.*
- *Your **evidence** definitely supports your thesis. I would recommend using actual **evidence** from the sources instead of simply summarizing. By including actual **evidence** you will be given more credibility, and it will just make your essay a lot stronger. You summarized the **evidence**, but you could have explained it a bit more. Explain the importance of the **evidence**. Why did you choose that specific **evidence** instead of another? Just find a way to explain it better.*

Comments of another weak-effect reviewer from Case 10 were typical in terms of length: their four form-focused comments were a lot longer (M = 152 words) than their eight content-focused comments (M = 74 words).

- *Very easily read. Overall well written. A few things here and there that could use revision. "To each sample of DNA, we added: Qiagen… " changes to "To each sample of DNA we added Qiagen…" "The next step was PCR (polymerase chain reaction)." should be reversed to "The next step was Polymerase Chain Reaction (PCR)… Citations used correctly where applicable, sentence formatting was very easy to read and understand. Many of your sentences start with "We" which felt pretty repetitive by the end. Maybe mix up a few more of those sentence beginnings.*
- *Methods are in chronological order from sample to extraction to PCR to sequencing. Transition sentences are underutilized here; a little more introduction to the following paragraphs and more of a summary sentence would be great!*

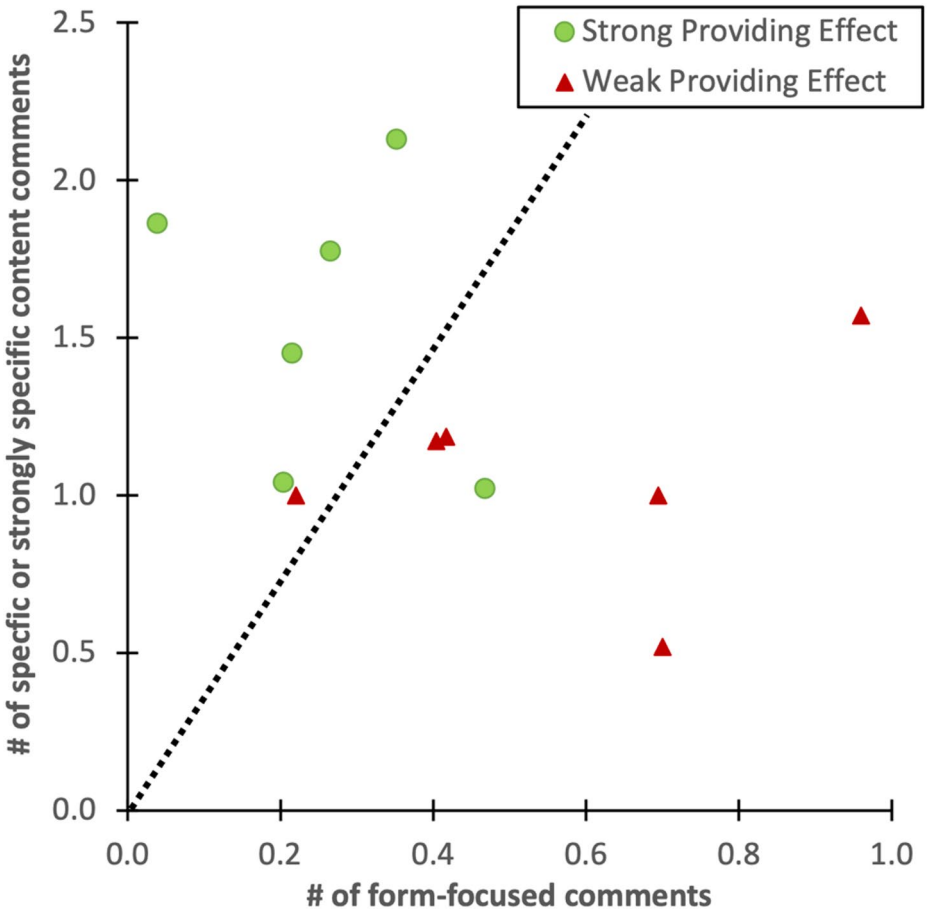

Fig. 1 The relationship between mean number of specific or strongly specific comments and the mean # of form-focused comments within the sampled long comments in the Strong and Weak providing effect cases. Dotted line defines a theoretical ratio that distinguishes most Strong and Weak cases

General comment differences

There was a trend for strong providing effect cases to include more general comments, but there was a lot of variation within each group, and the effect overall was not statistically significant (see 'others' in Table 4).

Integrating effects across content and form dimensions

Although there were large, statistically significant overall differences between strong and weak providing cases, we noted that there were some exception cases. To explore whether a combination of factors might explain the differences in learning outcomes more consistently, we graphed the interrelationships (see Fig. 1) between the largest content effect (the sum of *Specific* and *Strongly Specific* comments) and the largest form effect (overall number

of *Form* comments). On both dimensions in isolation, the general differences across were clear, but multiple overlap cases were also clear. Conceptually, the ratio of more specific content comments to form comments could also drive learner attention and thus learning differences.

By plotting a dividing line that was roughly three specific content comments per form comment, more strong and weak cases were clearly separated, with only two small deviation cases (i.e., Weak Case 11 just above the line and Strong Case 2 just below the line).

We then conducted a follow-up analysis of the two exception cases. Weak Case 11 above the line involved an engineering course in which the assignment was a graphic design of a motor. In this context, the coding of the content vs. form distinction in comments may not have been accurate. In particular, many of the comments referred to images in the assignment document, and these comments were coded as content-focused because to an outsider they seemed to involve technical knowledge. But on careful re-examination, these comments often actually involved discussions of violation of rules that therefore could be considered as form-focused. The two comments below are presented as examples of such comments, with ambiguous content vs. form aspects in bold. In other words, it was possible that this case had more form-focused comments, and thus should be further to the right, and then on the correct side of the classification line.

- Yes, the only suggestion I would make is that if you're not going to add small pictures to the Pugh chart to remind us of the design, **then consider at least titling the design proposals**. For example, proposal 1 could be " design 1:360 degree shaft" and proposal 3 could be "design 3: 2 shafts". This makes it so your reader doesn't have to scroll up to remind themselves of the general concept for each design.
- **The sketches are still clearly presented**. I'd still like to see a view of the final design that shows how the shaft and tray interface. Some alternate views of the final design (maybe one without the track) wouldn't have been beneficial. An engineer could still pick up the main concepts though.

Strong Case 2 below the line involved a very short writing task in an Anthropology course: a text-message posting selected from social media and a 150-word explanation of the post. Given that the writing assignment was so short (much shorter than the other assignments in the dataset, see Table 1), the number of content issue to comment upon was likely to be relatively small by comparison. It could be argued that the critical element for seeing strong learning was not the absolute number of specific comments; instead, it might be that the number of specific comments relative to the complexity of the writing was what drove higher levels of learning. Thus, in adjusting for assignment length, the relative level of specific comments for this assignment could be reclassified as higher, and then this case would side correctly on the strong providing effects side of the dividing line.

Discussion

Previous peer feedback research has shown benefits of providing feedback (Cho & MacArthur, 2011; Lundstrom & Baker, 2009; Zong et al., 2021). The current study aimed to explore how feedback comments in assignments with strong learning effects differ from those in

assignments with weak learning effects. To this end, the study investigated whether there were differences in terms of feedback focus and feedback specificity between strong and weak cases, especially when comment length was controlled. The study particularly focused on different levels of providers' behavioral and cognitive engagement behind those feedback features. The HLM analysis produced straightforward answers to the three research questions.

To clarify, the present study identifies differences between strong and weak cases; however, it does not establish that these differences account for the variation in learning outcomes. Yet, the contrast between high- and low-performing cases is inherently correlational, as participants were not randomly assigned to distinct learning conditions. A more rigorous test would require the systematic manual coding of thousands of comments produced by several hundred students across the twelve cases, followed by formal correlational analyses. Such an undertaking would be substantial and lies beyond the scope of the present study.

Learning from content comments

Regarding Hypothesis 1, the results demonstrated that the strong providing effect cases did not differ significantly in the number of content-focused comments from the weak cases. Instead, the difference was more in terms of comment specificity, as predicted by Hypotheses 2 and 3. This finding was not inconsistent with prior research that recognized the value of higher-level substance comments (Huisman et al., 2018; Patchan et al., 2016; Wu & Schunn, 2023). Yet, this finding provided additional evidence that relative learning gains were more likely to occur when providing *specific* rather than *superficial* comments on content issues.

From an engagement perspective, deep cognitive engagement involves more analytical thinking, critical evaluation, and a nuanced understanding of underlying concepts. Peer reviewers who merely attend to problems but without giving explanations or suggesting possible modifications are not deeply engaged with complex cognitive processes that drive learning (Van Popta et al., 2017; Zhang et al., 2023). Providing peer feedback of high specificity can challenge reviewers into a deeper meaning-making and knowledge-building process that might involve a number of complex cognitive processes that could support learning (van Popta et al., 2017). Therefore, we highlight the role of high-specificity content-focused feedback in learning from providing.

Learning from form comments

Regarding the second part of Hypothesis 1, as predicted, we found that form-focused revisions accounted for a larger proportion of comments among the weak providing effect cases. The weak-effect cases also tended to identify more error types and provide more direct error corrections, but this form of specificity did not seem to provide additional benefits to form-focused comments, as predicted by Hypothesis 3.

From an engagement theory perspective, responding to surface language issues, while important for refining written works and a valuable skill, may not demand the same level of deep cognitive engagement as tasks that involve complex analysis, synthesis, or creativity. Comments on surface language issues typically involve identifying and recognizing grammar mistakes, punctuation errors, or spelling mistakes. This is more about applying

well-established language rules rather than in-depth understanding or analysis. Therefore, surface language comments, no matter how detailed they are, may be generally associated with low cognitive engagement. This explanation would therefore suggest that form contents could be the basis of stronger learning if students approached the task of providing form-focused comments in a more cognitively engaging way. For example, if they had to evaluate patterns in form issues within a document or propose new strategies for addressing form-focused issues, this might involve higher levels of cognitive engagement and therefore learning.

As previously discussed in the literature, the division between content and form can be challenging to define and operationalize (Hyland, 2003). To this larger discussion about the dividing line, our study draws attention to this same issue within visual elements of documents (Case 11, a graphic design assignment). Visual elements can involve rules and conventions and therefore be considered issues of form. Our exception case suggests that form-focused visual comments are also a weaker learning opportunity. Future research will be required to investigate the nature of form-focused comments around visual aspects of documents to help more broadly refine the category distinctions and further uncover when and why such comments have more limited learning functions.

Integrating what and how for productive learning

Although we initially hypothesized in separate ways regarding feedback focus (the what of feedback) and feedback specificity (the how of feedback), our findings suggest an integrated perspective is more likely to be productive. In particular, rather than thinking about absolute amounts of focus and specificity, we argue the ratio of the two—highly specific content comments to form comment—is more useful for identifying situations that promote student learning from providing. Conceptually, these results could be understood from a limited cognitive architecture perspective. Students only have so much cognitive capacity and time (Sweller et al., 2011), and having students spend more of that cognitive capacity and time on feedback elements that involve higher cognitive engagement and sense making is more likely to lead to higher levels of learning.

General comments

We found that there was no significant difference between strong providing effect cases and weak effect cases in terms of general comments. These general comments were largely praise (for hard work and dedication) given both before and after the constructive or critical feedback. Providing sandwich feedback as such may help balance constructive criticism with positive reinforcement, making the feedback more palatable and motivating for the recipient (Nelson & Schunn, 2009). However, the literature is mixed on whether such feedback is actually useful for the recipient (Cheng et al., 2015; Hattie & Timperley 2007). In any case, the current study suggests such comments are neither helpful nor harmful for the provider.

Practical implications

Participants in the present study were drawn from relatively homogeneous populations—students at selective public universities in North America—and may therefore be characterized as relatively high-proficiency writers. However, individual proficiency levels were not available, as Peerceptiv does not require teachers or students to report such information. Consequently, the practical implications of these findings should be interpreted with caution, particularly with regard to their generalizability to lower-proficiency populations. Whether similar patterns would emerge among less proficient writers remains an open empirical question. Nevertheless, prior evidence suggests that both higher- and lower-proficiency writers are capable of providing accurate content-focused feedback when supported by feedback forms that direct attention to relevant issues (Wu & Schunn, 2020).

Juxtaposing the present findings on feedback features that most strongly support learning with evidence on the types of comments students typically produce highlights a need for additional instructional support. Research indicates that lower-proficiency writers tend to generate fewer revision suggestions and less specific feedback than their higher-proficiency peers (Zhang et al., 2023). Similarly, studies in ESL contexts have documented relatively low levels of cognitive engagement during peer review, with students focusing primarily on lower-order concerns (Fan & Xu, 2020; Yu et al., 2019). Encouragingly, even brief training interventions have been shown to help students shift their attention toward content-level issues and produce more specific, developmentally meaningful feedback (Cui et al., 2021; Ion et al., 2019).

A further implication arises in light of evolving sociocultural learning environments and advances in educational technology. On the student side, AI-supported tools could be designed to provide real-time guidance that promotes greater specificity and sustained focus on content in feedback comments. On the instructor side, enhanced interface features or "setup wizards" on the Peerceptiv system could assist in calibrating task design by guiding the balance between content- and form-focused feedback and by prompting greater specificity in peer review.

Limitations

We acknowledge that the current study has not considered all of the factors that likely shape how much students learn from providing feedback. We focused on features that were plausible candidates from the research literature, but other aspects of comments could also matter for learning from providing (e.g., the writing proficiency of student writers; the accuracy of the comments provided).

The current study also focused on explaining variation in near transfer learning. It could be that other factors are relevant to explain variation in document revision or far transfer learning. We suspect that deep cognitive engagement for the provider may be less important for document revision but potentially even more important for far transfer.

The assignments included in the present study were not fully controlled for text genre or assignment type. We acknowledge that the nature of peer comments may vary substantially across tasks—for example, between a brief written message and a full-length informative essay—and that such variation does not necessarily reflect differences in writer or reviewer

proficiency, but rather differences in task demands. We therefore agree that assignment type may have influenced the observed patterns and should be explicitly discussed as a potential limitation. At the same time, although the cases in the two groups were not perfectly matched, we did take steps to enhance comparability by considering text genre.

Another important limitation to consider relates to causality. The observed differences in comment features (e.g., focus and specificity) cannot be interpreted as causal explanations for the variation in learning outcomes. The results section asserts that specific feedback is correlated with higher learning outcomes but does not explore alternative explanations, such as the influence of students' baseline competencies or prior feedback experience. Without addressing these potential confounding variables, the study risks attributing causality where only correlation is evident, limiting the reliability of the conclusions drawn. The current study is quasi-experimental in design, sampling cases of higher and lower learning from feedback carefully matched on other dimensions. But, to fully establish causality, experiments in which students were assigned to providing more vs. less specific comments or form-focused comments would be helpful in testing the causality of the relationships revealed in the current study.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11251-026-09813-3.

Acknowledgements This work was partially supported by the China Scholarship Council (Grant Number 202207137053) and Dalian University of Foreign Languages (Grant Number 2022XJYB34).

Declarations

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

References

Chen, W., Liu, D., & Lin, C. (2023). Collaborative peer feedback in L2 writing: Affective, behavioral, cognitive, and social engagement. *Frontiers in Psychology*, *14*, 1078141.

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

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832.

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

Cui, Y., Schunn, C. D., Gai, X., Jiang, Y., & Wang, Z. (2021). Effects of trained peer vs. teacher feedback on EFL students' writing performance, self-efficacy, and internalization of motivation. *Frontiers in Psychology*, *12*, 788474.

Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing*, *50*, 100775.

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

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

Huisman, B., Saab, N., Van Den Broek, P., & Van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: A Meta-Analysis. *Assessment & Evaluation in Higher Education*, *44*(6), 863–880.
Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, *31*(2), 217–230.
Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101.
Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, *44*(1), 124–138.
Karim, K., & Nassaji, H. (2020). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research*, *24*(4), 519–539.
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.
Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, *41*(3), 525–536.
Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*(2), 193–211.
Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S., & K. Suen, H. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education, 41*(2), 245–264. https://doi.org/10.1080/02602938.2014.999746
Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance. *Assessing Writing*, *56*, 100725.
Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43.
Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375–401.
Ng, Z. Y., Tay, W. Y., & Cho, Y. H. (2015). Usefulness of peer comments for English language writing through Web-based peer assessment. In G. Chen, V. Kumar, R. Kinshuk, Huang, & S. C. Kong (Eds.), *Emerging Issues in Smart Learning* (pp. 385–392). Springer.
Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591–614. https://doi.org/10.1007/s11251-015-9353-x
Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, *108*(8), 1098–1120.
Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, *109*(8), 1049–1066.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory* (1st ed.). Springer New York.
Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, *21*(2), 217–235.
Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, *20*, 24–34.
Wu, Y., & Schunn, C. D. (2020). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology*, *60*, 101826.
Wu, Y., & Schunn, C. D. (2021). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science*, *49*(3), 365–394.
Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology*, *73*, 102160.
Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, *37*, 100484.
Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, *147*, 107857.
Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2019). Understanding student engagement with peer feedback on master's theses: A Macau study. *Assessment & Evaluation in Higher Education*, *44*(1), 50–65.
Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology*, *54*(4), 817–835.

Zhang, F., Schunn, C., Chen, S., Li, W., & Li, R. (2023). EFL student engagement with giving peer feedback in academic writing: A longitudinal study. *Journal of English for Academic Purposes*, *64*, 101255.
Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*(3), 372–386.
Zong, Z., Schunn, C. D., & Wang, Y. (2021). What aspects of online peer feedback robustly predict growth in students' task performance? *Computers in Human Behavior*, *124*, 106924.