Mu and Schunn *Int J Educ Technol High Educ* (2025) 22:4
https://doi.org/10.1186/s41239-025-00502-8

International Journal of Educational Technology in Higher Education

RESEARCH ARTICLE Open Access

The good, bad, and ugly of comment prompts: Effects on length and helpfulness of peer feedback

Huifeng Mu[1*] and Christian D. Schunn[2]

*Correspondence:
huifengm@pitt.edu

[1] School of Foreign Studies, Jiaxing University, No.899 Guangqiong Road, Jiaxing 314001, Zhejiang, China
[2] Learning Research and Development Center, University of Pittsburgh, 3420 Forbes Ave., Pittsburgh, PA 15260, USA

Abstract

Peer feedback can be highly effective for learning, but only when students give detailed and helpful feedback. Peer feedback systems often support student reviewers through instructor-generated comment prompts that include various scaffolding features. However, there is little research in the context of higher education on which features tend to be used in practice nor to which extent typical uses impact comment length and comment helpfulness. This study explored the relative frequencies of twelve specific features (divided into metacognitive, motivational, strategic, and conceptual scaffolds) that could be included as scaffolding comment prompts and their relationship to comment length and helpfulness. A large dataset from one online peer review system was used, which involved naturalistic course data from 281 courses at 61 institutions. The degree of presence of each feature was coded in the N = 2883 comment prompts in these courses. Since a given comment prompt often contained multiple features, statistical models were used to tease apart the unique relationship of each comment prompt feature with comment length and helpfulness. The metacognitive scaffolds of prompts for elaboration and setting expectations, and the motivational scaffolds of binary questions were positively associated with mean comment length. The strategic scaffolds of requests for strength identification and example were positively associated with mean comment helpfulness. Only the conceptual scaffold of subdimension descriptions were positively associated with both. Interestingly, instructors rarely included the most useful features in comment prompts. The effects of comment prompt features were larger for comment length than comment helpfulness. Practical implications for designing more effective comment prompts are discussed.

Keywords: Comment helpfulness, Comment length, Comment prompts, Peer feedback, Scaffolding

Introduction

Peer feedback is gaining increasing attention in research and practice (Double et al., 2019), especially in the context of higher education and with the support of online peer feedback systems (Dawson et al., 2024; Gao et al., 2023; Kerman et al., 2024; Little et al., 2024; Nicol et al., 2014; Zhang et al., 2024). Meta-analyses have established its value

in improving student outcomes, including improving task performance (Double et al., 2020; Huisman et al., 2019; Vuogan et al., Vuogan & Li, 2022) and student academic attitudes (Li et al., 2021). However, these same meta-analyses also found large heterogeneity of effects: sometimes the benefits are large and sometimes the benefits are small. Further, the heterogeneity was not well explained by simple contextual factors included in the meta-analyses like discipline, educational level, or type of object being evaluated, although training on peer feedback was the most important moderator (Li et al., 2020). A number of authors have drawn attention to concerns about the quality of the feedback that student produce: if the feedback is of very short or otherwise of low quality, it is likely of little value for both recipient and provider (Dong et al., 2023; Harks et al., 2014; Jin et al., 2022; Wu & Schunn, 2021a, 2022, 2023; Yu & Schunn, 2023; Zong et al., 2021a). Here we explore the hypothesis that guidance provided in the online reviewing task shapes the value of the peer feedback that is provided. In particular, we hypothesized that if the prompt for peer comments in the online reviewing form given to reviewers contains critical scaffolds (such as reminders of what aspects of the task to consider or how to provide more helpful feedback), then higher education students would produce longer and more helpful comments to their peers.

In research that examines peer feedback more generally, the comment prompts used to shape peer feedback have been referred to as question prompts (Jurkowski, 2018; Xun & Land, 2004), feedback prompts (Leijen, 2017), feedback provision prompts (Alqassab et al., 2018) as well as the formative assessment scripts (Alonso-Tapia & Panadero, 2010; Panadero et al., 2012, 2014; Peters et al., 2018). The specific prompts in that research were typically focused on relevant task criteria with specific questions that followed an expert model of approaching a task step-by-step, which facilitates the students in assessing their peer's work. Previous research on comment prompts have revealed that they can help students make more comments, better detect existing problems, and include good suggestions for revisions (Peters et al., 2018), and comment prompts were helpful for students to more deeply consider received feedback and to implement more revisions based upon received comments (Jurkowski, 2018).

However, there is very little work examining common practice: what kinds of comment prompts do higher education instructors typically include within online peer feedback systems? Studies often focus on the effects of researcher-designed comment prompts, which may be very different from actual practice in higher education, where instructors receive little pedagogical training (Gormally et al., 2014; Hamer et al., 2015; Morris et al., 2021; Paris, 2022). There is also relatively little prior research on the specific effects of comment prompt *details* on peer feedback. In particular, it remains unclear what forms of comment prompts facilitate students in producing long and helpful peer feedback in web-based systems. While prior studies have shown effects of different scaffolds as created by researchers, the ways in which instructors typically enact those scaffolds might have different effects. Therefore, the present research focused on further exploring the specific forms of comment prompts used in peer feedback and their underlying impacts on both length and helpfulness of peer feedback by answering:

> RQ 1. How commonly do higher education instructors provide different forms of scaffolding in comment prompts?

RQ 2. Which forms of scaffolding influence comment length?
RQ 3. Which forms of scaffolding influence comment helpfulness?

Literature review

Scaffolding in peer feedback

Scaffolding is an instructional approach in which an instructor provides support for a student to help them develop their skills and understanding and the support that is provided gradually decreases as the student becomes more capable and independent (Cook et al., 2020; Könings et al., 2019). This scaffolding can be conceptual (where to focus attention), strategic (approaches to consider), metacognitive (how to self-regulate), or motivational (why to keep going) (Belland, 2016). When applied to peer feedback, scaffolding suggests that students can benefit from receiving feedback from their peers, but even higher education students may need some guidance to provide high-quality feedback (Alemdag & Yildirim, 2022; Cui & Schunn, 2024). In this case, the instructor can act as the more knowledgeable individual and provide various forms of scaffolding (e.g., specific criteria for what makes good feedback and how to structure their feedback; Carson & Kavish, 2018). Applying scaffolds to peer feedback can help ensure that students are able to provide high-quality and useful feedback to their peers.

Theoretical classifications of comment prompts

Although there is very little research that examined the relative effects of different kinds of peer feedback prompts, some initial hypotheses can be made about potentially important features to include based upon studies that tested the benefits of a particular kind of comment prompt. Note that we see a comment prompt as a complex artifact that can contain multiple features to varying degrees, rather than simple categories to choose among. Further, we conceptualize these features as scaffolds for students (Cho & Schunn, 2007; Lee et al., 2021; Topping, 1998), meaning that they assist students in completing the peer feedback task when it is slightly beyond their own unassisted performance level. This assistance might involve noting more issues than they would without assistance or providing feedback in a way that is more complete or more helpful than they otherwise would. However, instructions and computerized learning environment features designed with scaffolding goals are not always successful (Kim et al., 2018; Zheng, 2016). First, scaffolds reminding students of what to do can be ineffective if students do not know how to do it (e.g., prompts that ask for constructive advice will not work if students cannot generate possible solutions to problems they identify) (Nguyen et al., 2016). Second, too much guidance can be overwhelming or demotivating to students (Kalyuga, 2011; Kirschner et al., 2018).

Given our first research question (i.e., what do instructors do?), we needed a categorization scheme that matched the range of what instructors do, rather than focus narrowly on only the specific scaffold types considered in a particular framework. Therefore, we inductively developed initial specific scaffold categories based upon an examination of existing practice. However, we refined and organized these categories based upon a theoretical framework regarding the different focus/functions of scaffolds. In particular, scaffolding embedded into online peer comment prompts was

classified into conceptual scaffolds, motivational scaffolds, metacognitive scaffolds, and strategic scaffolds (Belland, 2016; Belland, et al., 2013; Hannafin et al., 1999). Conceptual scaffolds give guidance on what conceptual issues to consider in the feedback (Belland, 2016; Sandoval & Reiser, 2004). Motivational scaffolds phrase feedback providing requests in ways that motivates participation (Belland, 2016; Tuckman, 2007; Wigfield & Eccles, 2000). Metacognitive scaffolds give guidance on how to self-regulate feedback giving (Belland, 2016; Cuevas et al., 2002). Finally, strategic scaffolds give guidance on how to provide feedback (Belland, 2016; Reiser et al., 2001). Here we review the specific scaffolds that could occur within each functional group, alongside prior research that could inform expectations of their effects.

Conceptual scaffolding

Prompting specific subdimensions. Comment prompts can focus student attention on particular aspects of an assignment or a project task, as a kind of conceptual scaffold. We term these specific subdimension prompts. Such prompts have been widely described in research as a guideline for student reviewers to make comments on specific global and local writing issues (Chang, 2016; Shvidko, 2020). Comment prompts naming specific global writing issues might focus on logic and support or organization, and comment prompts naming local writing issues might focus on spelling, grammar, or sentence structure (Leijen, 2017). Prior research on these comment prompts discovered that they helped students complete the reviewing task in a more expert-like way (Jurkowski, 2018; King, 2002; Nückles et al., 2009). They also helped students detect errors (Peters et al., 2018; Rietsche et al., 2022; Rotsaert et al., 2018), make more comments on their peer's performance (Alqassab et al., 2018; Gan & Hattie, 2014), better understand their peer's feedback and make better revisions (Jurkowski, 2018), and improved students' self-regulation and learning (Panadero et al., 2012).

Motivational scaffolding

Prompting with binary and open-ended questions. One way of motivating students is to increase their sense of agency (Deci & Ryan, 2012). Scaffolding students via questions may increase student agency relative to more direct instructions. Comment prompts can include both binary and open-ended questions. Binary question prompts consist of questions about the presence or absence of desired behaviors (i.e., could be answered with yes/no) and open-ended question prompts use questions requiring both in-depth and longer feedback (Bong & Park, 2020). Some research on their effects on length or helpfulness of peer feedback indicated that both binary question prompts and open-ended question prompts can be helpful (Bong & Park, 2020; Shiu et al., 2012). However, because open-ended question prompts can call for reflective responses (Ellegaard et al., 2018), they have been found to promote students' participation, cultivate students' logic thinking ability (Liu, 2019) and facilitate students in making more informative feedback, detecting more potential problems, and addressing constructive suggestions (Bong & Park, 2020).

Metacognitive scaffolding

In developing higher quality feedback, students may need scaffolds that allow them to judge for themselves whether they are providing effective feedback. This might be done by encouraging deeper commenting more generally or via more specific ways of characterizing higher quality feedback.

Prompting with elaboration requests. Comment prompts can also request that students elaborate on the feedback core components (i.e., strengths, weaknesses) through explanations, detailed descriptions, or discussions. Such elaboration prompts have been argued to activate students' schemata, help them articulate their thinking and reasoning, and prompt students to make more detailed explanations and discussions (Ge et al., 2005; King, 1992; King & Rosenshine, 1993; Nückles et al., 2009). Elaboration prompts are also thought be beneficial for students for constructing new knowledge by integrating what they have learnt before with details, examples, analogies and illustrations (Kobbe et al., 2007). Prior studies have found that requesting elaborated feedback was positively correlated with learning outcomes (Lee & Recker, 2021), improved students argumentative writing performance (Latifi et al., 2021), and produced more feedback (Peters et al., 2018).

Prompts that set expectations for high quality contributions or higher quality reviews. Beyond simply noting what elements are expected in the assignment (i.e., listing subdimensions), the comment prompts can also set expectations by providing more information about the functioning of those document elements (e.g., listing necessary components or what would constitute a strong contribution). Similarly, comment prompt can set expectations by describing what elements or features are required in a review (e.g., strengths, examples, or explanations), but they can also describe desired qualities of those elements (e.g., main strengths, salient examples, or clear explanations). Such prompts not only systematically instruct students on how to evaluate and comment on their peers' submissions, but they also remind and set expectations that should help students make more reliable and higher quality feedback, at least for online learners (Ertmer et al., 2010). However, little research has directly examined their relationship with the resulting length or helpfulness of peer feedback.

Strategic scaffolding

Another kind of scaffolding directs students on how to produce an effective comment in a procedural way (i.e., which commenting strategies to use).

Prompting with requests to identify problems, include examples, and include suggestions. A number of researchers have focused on specific cognitive features to include in a comment that improves its helpfulness: Identification of strengths and weaknesses, including suggestions for how to address identified problems, giving specific examples of general problems, and being specific about the location of problems. Emphasizing the inclusion of such elements in a comment prompt was designed to assist student reviewers to assess their peers' strengths and weaknesses on a wide range of task-related dimensions and generate suggestions (Dmoshinskaia et al., 2021; Wu & Schunn, 2021a, 2022). While some researchers include all three, others have focused specifically on identification (King, 1994) or suggestions. Feedback including specific location information was

preferred by comment recipients (Leijen, 2017; Nelson & Schunn, 2009; Patchan et al., 2016). Feedback including suggestions is preferred by students and they helped engage students in further thinking (Cowie, 2005). In tutorial systems, questions prompting for examples tends to produce longer responses (Graesser & Person, 1994). Many studies have found that helpful feedback involves detailed identification of problems and constructive advice on how to improve the detected problems (Gielen & De Wever, 2015; Huisman et al., 2017; Nelson & Schunn, 2009; Tseng & Tsai, 2007; Wu & Schunn, 2020, 2022; Zong et al., 2021b). Further, constructive suggestions in peer feedback are thought to benefit both comment providers and comment receivers (Deiglmayr, 2018; Wichmann et al., 2018). Listing examples was one of ways in producing more elaborated feedback (Kobbe et al., 2007). However, research has not directly investigated the effects on length or helpfulness of peer feedback of including prompts to identify strengths and weaknesses, suggestions for improvement, and including examples. Because both example and location require what and where the problems are, they were categorized into the same comment prompt feature in the current study.

Prompts that request summaries. Comment prompts can also require students to summarize the main points in the document they are reviewing. Little research has been focused on the effects of summary prompts on length or helpfulness of peer feedback. However, an indirect approach was used to investigate their impacts on perceived helpfulness of peer feedback (Cho & MacArthur, 2010; Cho & Schunn, 2007; Leijen, 2017; Nelson & Schunn, 2009; Wu & Schunn, 2020). One study suggested that peer feedback including a summary was perceived as especially helpful (Nelson & Schunn, 2009). Another indirect approach was to explore the effects of summaries in writing. For example, summary writing has been argued to promote students' analytic thinking and logic reasoning (Lamb & Etopio, 2019), enhance students' learning development, and foster students' critical thinking in science (Ferretti et al., 2009).

Prompts for specific numbers. Comment prompts can also require students to name a particular number of review content pieces. Some peer feedback systems include the possibility of requiring a minimum number of words in a comment. But the comment prompt itself can also provide guidance on expectations for a specific number of some comment aspect (e.g., number of criticisms or number of suggestions). However, little is known about direct relationship of prompts listing a specific number and length on helpfulness of the resulting peer feedback even though it is logical that having more examples or listed problems would make longer comments (Neubaum et al., 2014). In terms of helpfulness, previous research has observed that longer comments were perceived as more helpful (Jin et al., 2022; Patchan et al., 2016; Wu & Schunn, 2022; Zong et al., 2021a, 2021b, 2022). More broadly, text length is generally correlated with text quality (Crossley, 2020; Fleckenstein et al., 2020; MacArthur et al., 2019). However, it is not clear that listing specific numbers of elements to include in a comment prompt will result in more feedback or more helpful feedback.

Literature review summary

Overall, researchers have created and tested the impacts of many kinds of peer feedback prompts, and, in isolation, there is some support for including many different features in a comment prompt. However, it is likely that comment prompts that include all

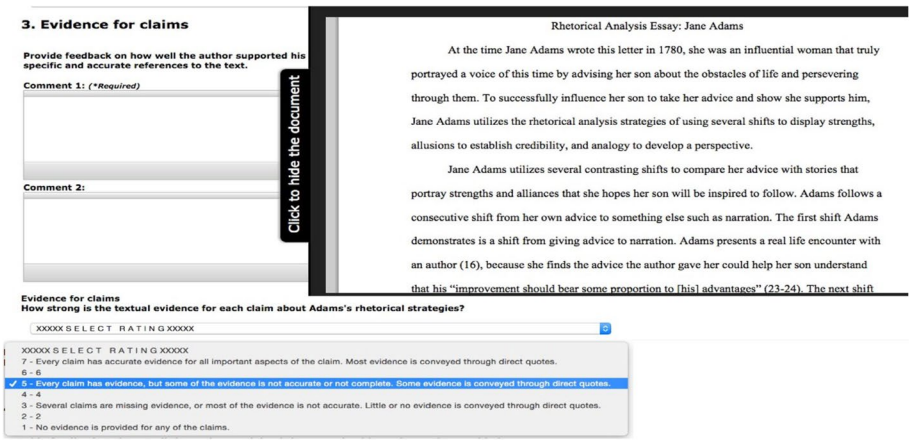

Fig. 1 The reviewing interface of Peerceptiv during the time of collected data

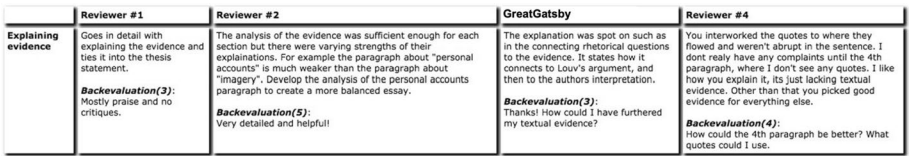

Fig. 2 The back-evaluation interface of Peerceptiv during the time of collected data

those elements would be overwhelming for students and burdensome for instructors to create. Further, no research has looked at the relative value gained by adding different features (e.g., does noting subdimensions within a comment prompt matter as much as prompting for strengths and weaknesses?). Moreover, little is known about what higher education instructors tend to do in typical practice, and whether the ways in which they include the recommended comment prompt features also improve student feedback.

Method

Dataset

The dataset was produced by a script applied to instructor-created peer feedback assignments and accompanying student feedback comments given to peers via the online peer review system Peerceptiv, initially called SWoRD (Schunn, 2016). Most relevant to the current research, in the reviewing interface, students were given comment prompts with textboxes for students as reviewers to provide comments (Fig. 1 left). This system contains a number of features that support best practice in scaffolding effective peer feedback, similar to a number of other web-based peer feedback systems like Eduflow, FeedbackFruits, and EliReview. In particular, multiple reviewers were assigned to each document, reviews were made anonymously to authors to improve honesty of feedback, and specific prompts guided the content of requested feedback. Most relevant to the current study, authors evaluated the helpfulness of the feedback they received, on a 1-to-5 scale called back-evaluations alongside optional comments (see Fig. 2), to produce a grading incentive for higher quality feedback.

The script downloaded all assignments for every available course and every commenting prompt in each assignment. The script calculated the mean back-evaluation scores

and mean comment length provided across students for the given comment prompt in the given assignment in the given course. These two mean scores constituted the two main outcome variables.

The script was run in 2015, collecting data from courses in universities around the world that took place over the window from 2010 to 2015, reflecting a period during which user agreements allowed for inclusion of all of this data in research. In addition, in later years, instructors had the ability to copy from a large library of other instructor's assignments and comment prompts, whereas there was a very small library during the studied period, providing a better estimate of what instructors tend to produce on their own.

Only courses with at least 25 students were included in analyses to produce sufficient data on comment length and helpfulness. These courses produced 2999 comment prompts. 90 comment prompts were written in languages other than English, and these were excluded since coding these comments required additional expertise to code but there were relatively few of them. In addition, another 26 prompts were deleted because they were rating prompts. The remaining 2883 comment prompts (representing 281 courses) were then used in systematic coding and analysis.

Measures

Frequency of comment prompt scaffolds. The 12 different comment prompt scaffolds described in the literature review were coded within each comment prompt (see Appendix A for exact coding definitions and examples). Since some comment prompts repeated across assignments or courses, only 1,075 unique comment prompts needed to be coded by hand, and then formulas were used to copy coding values to all instances in the dataset. Each scaffold could occur multiple times within a given prompt. Using the coding manual, the coder marked each scaffold occurrence within the prompt.

Reliability was tested using a second coder who coded 100 randomly selected comment prompts. Reliability was assessed by the correlation between the number of a given scaffold found within each of the double-coded prompts. A correlation of 0.6 or higher between coders was obtained for all but four of the scaffolds. Coders met to discuss disagreement cases, and the coding manual was refined. Reliability for the four remaining scaffolds was assessed on another 40 randomly selected comment prompts, producing reliabilities of at least 0.89. All but one of the full set of 12 scaffolds had reliabilities above 0.88 (see Appendix A), indicating very strong coding reliability.

Outcome measures: mean comment length and comment helpfulness. Across all comments produced for each comment prompt within a given assignment, the mean comment length across students (i.e., the mean number of words produced in a typical review from one student evaluating one other student's work on just this one comment prompt) was obtained from the downloaded source data. These mean comment length values varied widely, from as low as nine words to as many as over 1,000 words on average (see Table 1). The reviewing interface did not require a minimum number of words in a comment.

Similar to comment length, mean comment helpfulness across students for a given comment was obtained from the downloaded source data. Such comment helpfulness ratings have been examined in a number of studies on peer feedback (see recent

Table 1 Descriptive statistics for outcome measures (N = 2883)

Measures	Length	Helpfulness
Mean	154	4.18
Standard deviation	99	0.43
Maximum	1,114	5.00
Minimum	9	2.72
Correlation with length	–	0.26**

** $p < 0.01$ level

scoping review by Misiejuk & Wasson, 2021). The rating scale used by students was on a 1-to-5 scale, but the observed mean values (see Table 1) ranged from 2.7 to the maximum possible value, 5.0, with a relatively high overall mean of 4.2. The two outcome measures were only modestly correlated with one another ($r = 0.26$, $p < 0.01$) and thus were treated as separated outcomes (see Table 1).

Data analysis

All data analysis was performed using SPSS version 29. Means, standard deviations, and percentage of occurrences that were zero (i.e., the scaffold was not included at all) were examined to uncover the relative frequency of comment prompt scaffolds in typical instructor practice.

Then linear correlations among comment prompt scaffolds were examined to identify whether multiple regression would be needed to separate out the effects of each comment prompt scaffold as well as identify potential problems of multi-collinearity in the case of multiple regressions models. An Exploratory Factor Analysis (with Promax rotation) was conducted to test whether the comment prompts scaffolds could be combined into a smaller number of factors (especially with a scaffolding category), and this analysis established that there was relatively weak factor structure, and it was better to leave the comment prompts as separable, independent scaffolds.

Further, visual inspection of the relationship between scaffolds and outcomes suggested that there were often non-linear and sometimes curvilinear effects. Therefore, all 12 raw scaffold count variables were converted into categorical variables (e.g., 0, 1 + or 0, 1–2, 3 +) by examining frequency histograms, to ensure there were sufficient power by category overall.

To formally test the relationship of comment prompt features with comment length and helpfulness, a series of ANOVA models were created in a sequential model-building fashion, separately for the two outcome variables. First, simple effects were tested of each comment prompt categorical variable in insolation. Second, a full model was tested that included all comment prompt categorical variables as simultaneous predictors. Third, a final model was tested in which only statistically significant comment prompt predictors were retained. The statistical significance and direction of effects were noted for each predictor in each model. We also examined whether the results held when controls for discipline of the course was included, which also indirectly controlled for type of assignment.

The effects of each of the statistically significant comment prompt scaffolds in the final model were then examined to establish what comment prompt scaffold levels were positive (or negative) factors influencing comment length or comment helpfulness. These patterns were then compared with the relative frequency of each level to determine the extent to which instructors tended to create optimal scaffolds.

Results

RQ1: How commonly do higher education instructors provide different forms of scaffolding in comment prompts?

The detailed information regarding the percentages of these instructor-provided comment prompts is presented in Fig. 3, organized by type of scaffolding. Only conceptual scaffolding was mostly included in a majority of comment prompts. At least one form of motivational, metacognitive, and strategic scaffolds occurred in nearly 50% of prompts. A number of specific scaffolds appeared in fewer than 25% of prompts.

The basic descriptive statistics for the raw comment feature variables are presented in Appendix B. Many of the features were most commonly absent, and summary was very rarely included. Indeed, only subdimension and binary question were included in a majority of comment prompts. On the other hand, when a scaffold was included in a comment prompt, for all but one of the scaffolds (summary) there were sometimes had multiple instances within a comment prompt—as often as 31 times for subdimension or between 7 and 10 times for open questions, binary questions, weakness identification, and expectation.

Appendix B also presents the linear intercorrelations among the comment prompt features. In five cases (8%), the raw counts were strongly correlated with one another (i.e.,

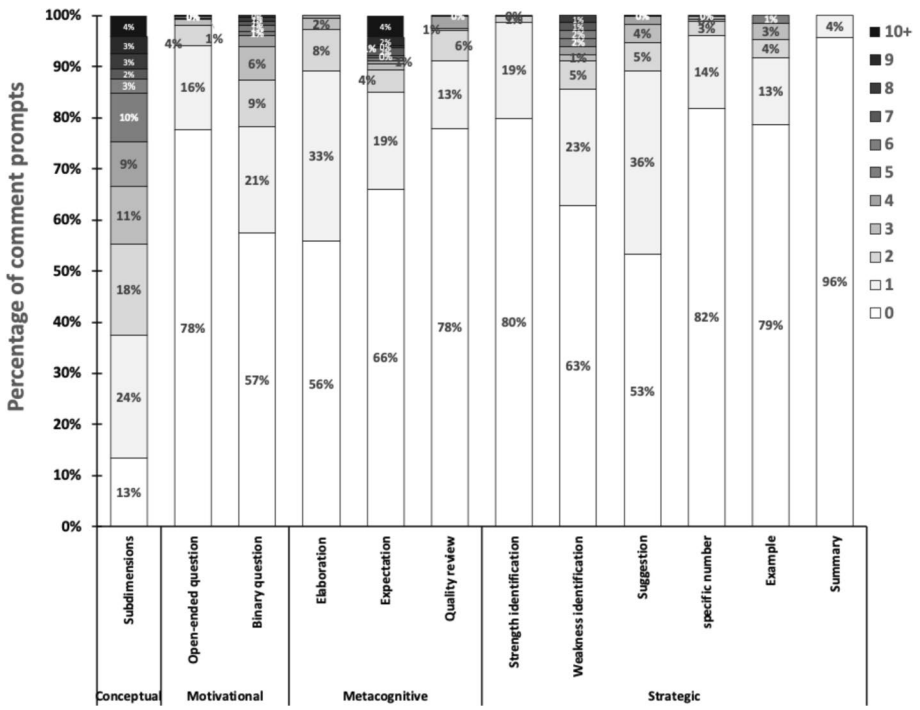

Fig. 3 Frequency of scaffolding by type

r > 0.65). Another 11 cases (17%) involved moderate correlations (i.e., between 0.35 and 0.65). 12 cases (18%) involved weak correlations (i.e., between 0.2 and 0.35) correlations. The remaining 38 cases (58%) involved very weak or even slightly negative correlations. Thus, many of the comment prompt features were independent of one another, but a few cases were sufficiently correlated with one another that multiple regression techniques were required to tease apart the unique contributions of each prompt feature on comment length and comment helpfulness. To address the few cases of very high intercorrelations as well as to address several non-linear relationships between intensity of scaffolding and length or helpfulness, three or four level categories were created for each comment prompt feature based upon visual inspection of frequency histograms.

RQ 2: Which forms of scaffolding influence comment length?

Table 2 presents the ANOVA findings across the three tested models for comment length. In some cases, the addition of covariates revealed relationships that were otherwise not statistically significant and in other cases, simple bivariate relationships became non-significant or reversed direction when adding covariates.

The variables are organized in Table 2 by the pattern of effects in the final model; the direction of the effects focuses on the largest effects, as revealed in the next section. Eight of the comment prompt features were statistically significant predictors of comment length, almost always at $p < 0.001$.

Figure 4 shows descriptive statistics for the specific scaffolds showing significant relationships between scaffold frequency and comment length, grouped by scaffold type. In terms of conceptual scaffolds, comment prompting subdimensions ($p < 0.001$) were positively related with comment length. In terms of motivational scaffolds, binary questions were positively related to comment length ($p < 0.001$), with a large difference when there were 4 or more binary questions, but open-ended question, showed a smaller positive effect ($p = 0.044$) at the low-end of the scale, and actually a negative effect when 3 or more open-ended questions were included. In terms of metacognitive scaffolds, both expectation ($p < 0.001$) and elaboration ($p < 0.001$) prompts were positively related with the comment length. Interestingly, two or more expectation prompts were needed to produce a benefit whereas one elaboration prompt sufficed. In terms of strategic scaffolds, none of them were positively related with comment length. Three showed negative relationships with length: suggestion ($p < 0.001$), specific number ($p < 0.001$), and example ($p < 0.001$).

To understand relative effect sizes, Fig. 5 shows the mean estimated effects for all scaffolds with significant effects on comment length, focused on the largest category differences. Metacognitive and conceptual scaffolds had moderate and consistent effects. Motivational scaffolding had inconsistent and moderate to large effects. Finally, strategic scaffolds had small-to-moderate negative effects.

RQ 3: which forms of scaffolding influence comment helpfulness?

Table 2 also presents the ANOVA findings across the three models for comment helpfulness. Seven of the comment prompt features were statistically significant predictors of comment helpfulness, but sometimes at more modest p values. Figure 6 shows descriptive statistics for the specific scaffolds showing significant relationships between scaffold

Table 2 Statistical significance and direction of effects in predicting mean comment length and mean comment helpfulness for each comment prompt feature predictor (organized by pattern of effects) across statistical models (single predictor, full modeling, final model with only significant predictors)

Pattern Predictor	Comment length			Comment helpfulness		
	Simple effects	Full model	Final model	Simple effects	Full model	Final model
Consistent effect on both						
Conceptual scaffolding						
Subdimension	+++	+++	+++	+++	+++	+++
Motivational scaffolding						
Open-ended question	ns	ns	–	– – –	– – –	–
Opposing effects						
Metacognitive scaffolding						
Elaboration	+++	+++	+++	+++	– – –	– – –
Strategic scaffolding						
Example	– – –	– – –	– – –	+++	ns	++
Effects on length						
Motivational scaffolding						
Binary question	+++	+++	+++	+++	– – –	
Metacognitive scaffolding						
Expectation	+++	+++	+++	+++	+++	
Strategic scaffolding						
Suggestion	ns	– – –	– – –	+++	– – –	
Specific number	– –	– – –	– – –	ns	– – –	
Effects on helpfulness						
Metacognitive scaffolding						
Quality review	++	ns		+++	+	– – –
Strategic scaffolding						
Summary	+	+		++	–	– – –
Strengths identification	+++	ns		+++	+++	++
No effects						
Strategic scaffolding						
Weaknesses identification	+++	++		+++	+++	

Omitted cases means that the variables were not included in the model. ns = not significant. For positive effects, + = $p < 0.05$, ++ = $p < 0.01$, +++ = $p < 0.001$. For negative cases, – = $p < 0.05$, – – = $p < 0.01$, – – – = $p < 0.001$

frequency and comment helpfulness, grouped by scaffold type. In terms of conceptual scaffolds, prompting subdimensions was positively related with the comment helpfulness ($p < 0.001$). In terms of motivational scaffolds, only open-ended questions was significantly related with comment helpfulness ($p = 0.019$) and with a negative relationship. In terms of metacognitive scaffolds, both elaboration ($p < 0.001$) and quality review ($p < 0.001$) were negatively related with comment helpfulness. In terms of strategic scaffolds, strength identification ($p = 0.004$) and example requests ($p < 0.001$) had a positive relationship with helpfulness but and summary requests ($p < 0.001$) were negatively related to comment helpfulness.

To show relative effect sizes, Fig. 7 shows the largest category contrast for all scaffolds with significant relations with comment helpfulness. Conceptual scaffolds had a small

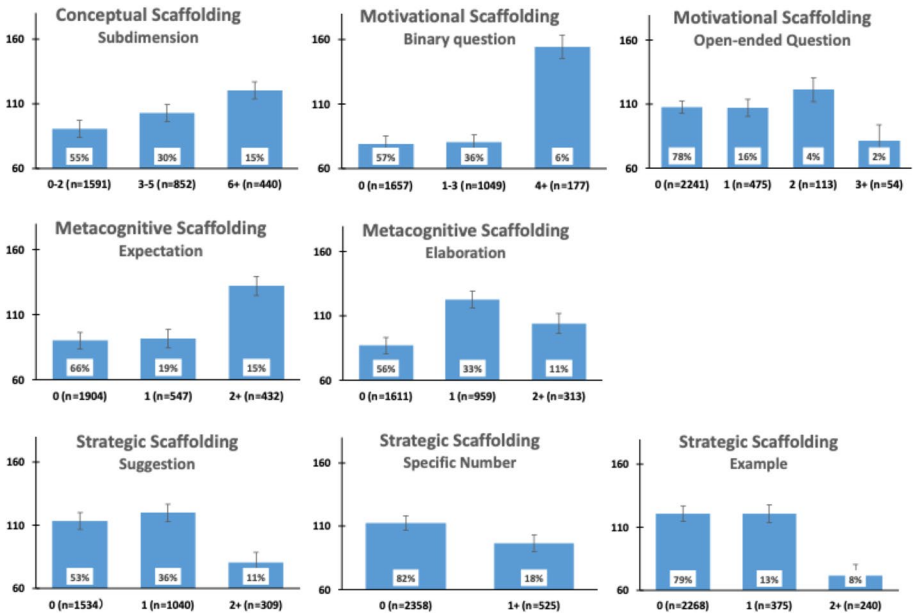

Fig. 4 Marginal mean comment length for each statistically significant comment prompt feature, controlling for effects of other comment prompt features. Error bars represent standard errors, and the relative frequency of each comment prompt feature category is shown within each bar

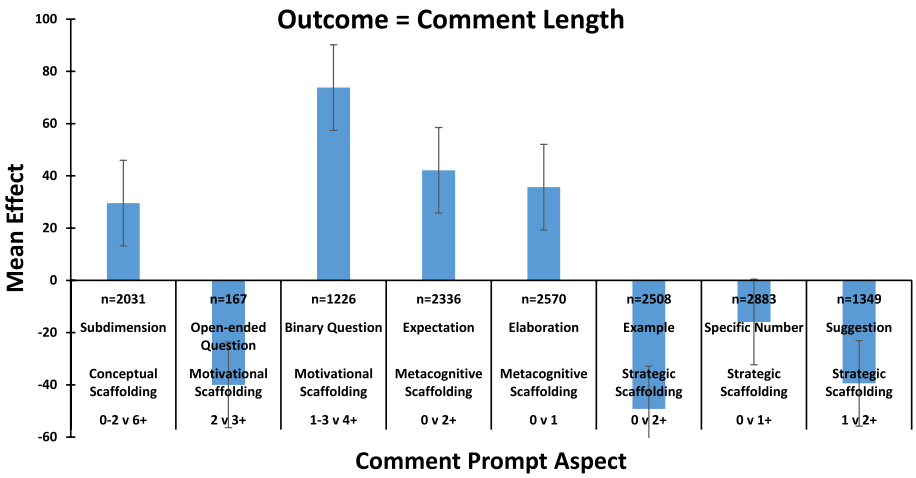

Fig. 5 Mean estimated effect on comment length for each of the significant comment prompt aspects in the final models

positive effect, as did two of the three strategic scaffolds. However, the third strategic scaffold had a negative effect. Motivational and metacognitive scaffolds had small negative effects.

General discussion

RQ1. How commonly do higher education instructors provide different forms of scaffolding in comment prompts?

In general, instructors rarely included these comment prompt scaffolds, and almost no instructor included at the prompt scaffolds at the intensity levels that are especially

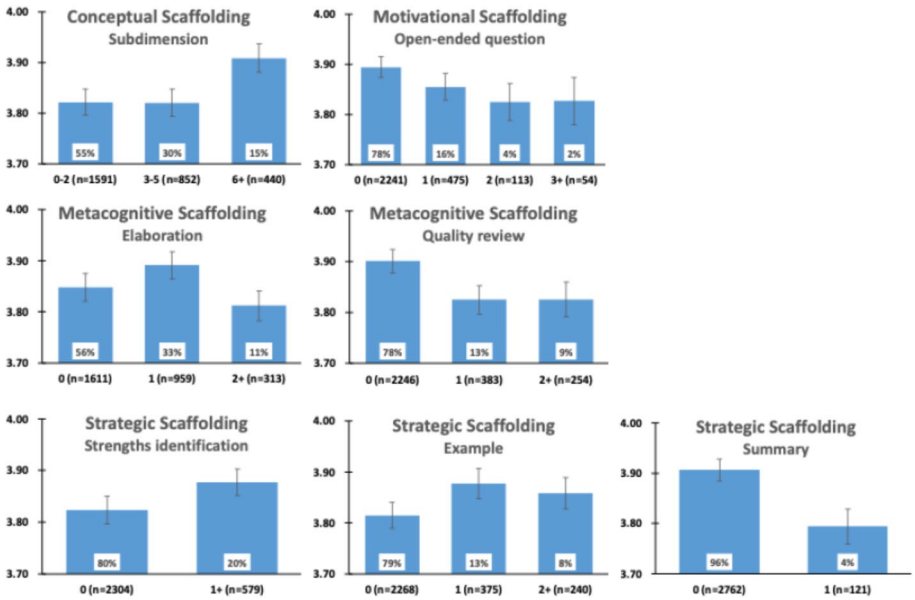

Fig. 6 Marginal mean comment helpfulness for each statistically significant comment prompt feature, controlling for effects of other comment prompt features. Error bars represent standard errors, and the relative frequency of each comment prompt feature category is shown within each bar

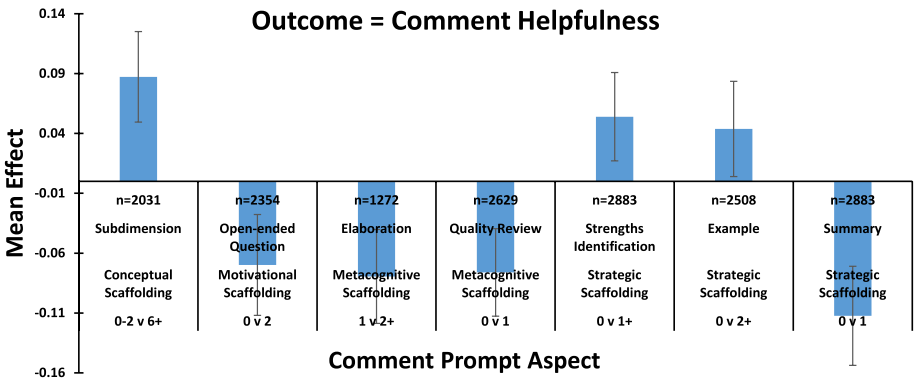

Fig. 7 Mean estimated effect on helpfulness for each of the significant comment prompt aspects in the final models

effective. It suggests that design of effective comment prompts is likely not an intuitive element of teaching.

A few comment prompt scaffolds were included more often. Only subdimension, a kind of conceptual scaffold (88%), and binary question, a kind of motivational scaffold (51%), were relatively commonly used as scaffolds when instructors designed comment prompts. Since the conceptual scaffold of subdimensions is closely tied to the basic design of the assignment, these might be especially salient or important to instructors. By contrast, the motivational scaffolds of binary questions can be taken as an easy way of prompting students on these elements; that is, perhaps it was common because it was easy to generate.

While it was reasonable for instructors to exclude prompt scaffolds that generally had negative associations, in general instructors did not seem to include the more useful comment prompt scaffolds. For example, the metacognitive scaffolds of elaboration and expectation and the strategic scaffolds of strengths identification were effective comment prompt scaffolds to improve comment length or helpfulness, but were rarely provided. Perhaps instructors are not receiving feedback on the effectiveness of their design considerations that naturally produces a shift to more effective prompt designs. Alternative, if instructors are rarely including comment prompt features, the systems they use do not provide information that would help instructors judge the effectiveness of the prompts.

No prior research had examined what scaffolds instructors typically include. This study revealed that, overall, instructors tend to rarely support students with comment prompt scaffolds in their teaching, and they do not support as many comment prompt scaffolds that have been found to be especially effective as possible even when they do.

RQ 2. Which forms of scaffolding influence comment length? Three of the four features that had a positive relationship with comment length (see Table 2) generally drew attention to various aspects of the assignment that were important (Alqassab et al., 2018; Gan et al., 2014). The conceptual scaffold of subdimension ($p < 0.001$) drew attention to the general elements and the metacognitive scaffolds of expectations ($p < 0.001$) reminded reviewers of the qualities that were most important within those elements. Conceptually, the motivational scaffolds of binary questions ($p < 0.001$) (Bong & Park, 2020; Shiu et al., 2012) can be thought of as a simple way of quickly probing about the important elements, either generally towards subdimensions (e.g., did the author attend to…) or more specifically towards expectations ($p < 0.001$) (e.g., did the author successfully …). The fourth element, elaboration ($p < 0.001$), as a strategic scaffold, asks reviewers to add important details they might not naturally include (Ge et al., 2005; Nückles et al., 2009; Peters et al., 2018).

Natural inclinations for the strategic scaffolds, such as what to include, might explain why strengths, weaknesses, suggestions, and examples did not increase comment length (see Table 2). Previous research showed that three of the four (weaknesses identification, suggestion, and example) were indeed common elements of peer feedback. Weaknesses were identified in between 24 and 45% of comments (Nelson & Schunn, 2009; Patchan et al., 2016; Wu & Schunn, 2020); general suggestions or specific solutions were identified in between 26 and 55% of comments (Jin et al., 2022; Nelson & Schunn, 2009; Patchan et al., 2018; Wu & Schunn, 2021b); and examples were identified in between 25 and 42% of comments (Nelson & Schunn, 2009; Patchan et al., 2016, 2018). However, it is especially interesting that suggestions and examples actually were associated with decreases in comment length. Perhaps by drawing attention to what reviewers would naturally do, students were less likely to go beyond those elements.

Overall, consistent with some of the prior research, comment prompts of subdimension, elaboration, binary question, and expectation were found to be helpful in guiding students to make longer peer feedback. Inconsistent with prior research, a number of other comment prompts tended not to be helpful, perhaps because students did not often need those scaffolds.

RQ 3. Which forms of scaffolding influence comment helpfulness? It is interesting that there was relatively little overlap between what scaffolds were associated with longer comments and what scaffolds were associated with more helpful comments. This finding is both novel and important in that it reveals that length cannot be assumed to be conceptually similar to helpfulness as an outcome. The conceptual scaffold of subdimension ($p < 0.001$) was the one element in common: authors also find comments more helpful when reviewers are directed towards the many elements needed to be included in the submission.

However, in terms of the scaffolds to comment in certain ways, the pattern was essentially the opposite (see Table 2): the metacognitive scaffolds of elaboration were negative and examples were positive. On the one hand, it may be that elaborations that reviewers provided were not accurate or simply stated in unhelpful ways (Gao et al., 2023). On the other hand, the strategic scaffolds like suggesting to include examples ($p < 0.001$) may lead to comments that are perceived as especially easy to action since the authors are told where in the document to make repairs (Leijen, 2017; Nelson & Schunn, 2009; Patchan et al., 2016). Further, the positive association of strengths identification ($p = 0.004$) with comment helpfulness likely relates to the motivational aspect of the produced feedback (rather than the motivational effects on the provider): students often find overly negative feedback as demotivational (Ellis, 2013; Hasan & Rezaul Karim, 2019; Hill et al., 2021).

Implications of the study

Prior research has found that students claim that "suggestions for improvement" (Viberg et al., 2024) were not an effective comment prompt because they think that these suggestions provided by their peers were not a significant part of their learning. Instructors are encouraged to provide students with more targeted comment prompts they think are beneficial for their learning (Jiang & Ironsi, 2024). Some elements of scaffolding functions may be particularly important in the context of higher education, where students tend to produce short and less helpful comments.

First, instructors are especially encouraged to explicitly note many assignment-specific dimensions in their comment prompts (e.g., as many as six specific subdimensions and not just one or two). An example of such a comment prompt is as follows: *Provide feedback on the student's controlled, sophisticated use of language: address its vocabulary (diction), syntax and grammar. Be specific about how the writer could improve his or her control of language and structure, and provide suggestions for improvements.*

Second, instructors are also encouraged to include the metacognitive scaffolds of elaboration and expectation, and the motivational scaffolds of binary questions in their comment prompts. In particular, the instructors should provide at least four binary questions (e.g., *Comment on the author's characters and plot structure. Was there anything you particularly liked about the main character? Any way the main character could have been improved? Were any scenes particularly effective? Could any scenes be improved upon, or removed? Be specific and helpful in your comments.*). In addition, they could provide at least two expectations (e.g., *Provide feedback on how well the author supported his or her argument with evidence. If any evidence was inaccurate, or if any of the author's points lacked evidence to back them up, make sure to point these out specifically and recommend how the author could improve.*), but preferably no more than

one elaboration request (e.g. *Identify the main strengths and weaknesses of the document in terms of the reasoning/support that was provided for the main claims or arguments. Be specific. Provide clear suggestions for improvement.*) to encourage students to make longer comments.

Third, instructors are discouraged from using some of the motivational scaffold and strategic scaffolds, particularly in heavy doses for open-ended questions (e.g., *How well did the author describe the specific utility of the technology? What is the broader significance of this technology and how does it advance our understanding of the relevant field?*), suggestions (e.g., *Provide clear suggestions for improvement.*), and examples (e.g., *Please provide suggestions for improvement and include at least one specific example of an error.*), or at all, in the case of specific number. Regarding specific numbers, it may be that minimal values, in particular, were problematic, and future research is needed to better understand what should be avoided.

Fourth, instructors should include the strategic scaffolds of strengths identification and examples. Note, however, that because of the negative effects of multiple example scaffolds on comment length, including just one example scaffold in the prompt, in particular, is what is recommended. For strengths identification, even just one such scaffold appears to be sufficient (e.g., *Make sure to explain what was specifically done well*).

Fifth, instructors are discouraged from including the scaffolds of open-ended questions, quality review, elaboration, and summary. Including just one elaboration scaffold appears to be safe, but the other elements appear to reduce comment helpfulness when included at all (e.g., *Provide a short but clear suggestion for making the question even more unambiguously correct.*)

Overall, several elements of comment prompts can help students make or receive longer and more helpful peer feedback. Instructors are encouraged to include these comment prompts as more effective scaffolds for the students in their teaching.

Caveats and future directions

While this research identified which comment prompt features (and at which levels) were robustly correlated with comment length or helpfulness, and which ones were not, such statistical associations are inherently correlation evidence and thus do not directly prove that they caused changes in comment length or helpfulness. This research design did include many controls in the regressions such that the likelihood of third-variable confounds is reduced. However, future studies in which the presence of comment prompt scaffolds was experimentally manipulated would be useful to more directly test the causal effects. The current research did rule out many kinds of comment prompt scaffolds that are unlikely to produce large effects, and such ruling out will be helpful in narrowing the focus of what should be tested in future experiments.

Second, the generality of the findings needs to be considered. Although drawn from a large dataset and explicitly examined from the perspective of generality across contexts, the findings might not generalize to contexts that were not studied. Saliently, the dataset was relatively sparse for courses in some non-English speaking countries, and disciplines beyond STEM or English. Future studies might specifically focus only on English-speaking or only non-English speaking countries, or only on STEM courses or only English courses.

Finally, all the data were collected from one online peer review system. Although this system contains features that are now relatively common in online peer feedback systems, there are also many peer feedback methods/systems that do not contain these features or use a very different medium (e.g., face-to-face comments or audio-recorded comments). Future research is necessary to collect more data from a variety of online peer review systems, especially ones that have substantially different features that might differentially scaffold comment length (e.g., having minimum length requirements) or comment helpfulness (e.g., not having accountability for more helpful comments).

Conclusion

This research specifically focuses on specific prompt features organized by scaffolding functions and their relationship with both comment length and comment helpfulness based on scaffolding theory. Twelve comment prompt features, representing four scaffolding functions, were carefully examined. The statistical models uncovered which comment prompts were associated with longer and/or more helpful comments, as well as which comment prompts were actually associated with shorter and/or less helpful comments. Relatively few comment prompt features showed robust positive relationships, and these comment prompt features were rarely included in typical practice. This research also reveals where peer feedback needs to include guidance for instructors to help them design more effective comment prompts.

Appendix A

Within each scaffolding type, definitions for the 12 comment prompt features, along with examples and with coding reliability correlation coefficients. Examples often contain multiple instances of the given scaffold, and the count is indicated in [].

Scaffold	Definition	Examples of Prompts (with key elements in bold)
Conceptual scaffolding		
Subdimension (r = 0.89)	Focuses on specific aspects of a dimension or consists of detailed aspects of a dimension	Give the student a comment about the quality of their summary. Comments could be about any of the following: The accuracy of the summary; Any grammatical errors you found; Unnecessary information; Their summary gave you a new perspective on your own summary. [4]
Motivational scaffolding		
Open-ended question (r = 0.91)	Questions about the reviewed object that require an open-ended response	What was your favorite part of their lab report? How well do the writing style and vocabulary in the paper fit what you expect for this kind of writing? [2]
Binary question (r = 0.97)	Questions about the reviewed object that can be answered yes or no	Did the level of detail and images used in presenting the chosen design adequately convey the design intent? Did the models emphasize the important parts of the design? Do you feel that another engineer could pick up the concept based on this presentation? [3]
Metacognitive scaffolding		

Scaffold	Definition	Examples of Prompts (with key elements in bold)
Elaboration (r = 0.94)	Specifies elaboration of feedback core components (e.g., strengths, weaknesses) through explanations, detailed descriptions, or discussions. Excludes requests for suggestions, examples, and locations	If any of the graphs or tables were unclear, please describe what titles/ labels could be added to help with the understanding. Also, explain what details could be included in the text so that the graphs or tables would be clearer. If the author did not answer all of the questions in the lab manual accurately, please indicate which questions were not answered and describe any inaccuracies. [3]
Expectation (r = 0.90)	Specifies what elements the reviewed object should contain (e.g., a list of necessary components) or what constitutes specific parts in the reviewed object (e.g., the character of a necessary component)	The conclusion should build on the report's introduction to explain how the results address a larger biological problem. The purpose of the Introduction section is to describe the rationale behind the experiment. [2]
Quality review (r = 0.96)	Describes aspects of a high-quality review (not a high-quality reviewed object)	Describe the main strengths and the main weaknesses in the organization and structure of the essay. Be specific and provide clear suggestions for improvement. [3]

Scaffold	Definition	Examples of Prompts (with key elements in bold)
Strategic scaffolding		
Specific number (r = 0.98)	Names a particular number of review content pieces	Provide 3 examples of descriptive language from the writer's personal statement. Give 2 examples of where the writer could use more descriptive language. [2]
Example/ location (r = 0.89)	Requires specific examples or where the issues are found are in the reviewed object	Did the student's writing convey a clear understanding of the assignment? Give specific examples [1]
Suggestion/ advice (r = 0.89)	The prompt requires specific suggestions or advice for how to improve a reviewed object	Provide specific feedback on how the writer can improve their introduction. [1]
Identification of Strengths (r = 0.63) Weaknesses (r = 0.97)	The prompt requires to identify strength/what's good or weakness/ what's bad or both in a reviewed object	Identify the main strengths and weaknesses of the introduction in terms of the reasoning/support that was provided for the main claims or arguments. [1 strength] [1 weakness]
Summary (r = 0.89)	Requires a summary of reviewed object	Write a short summary of the paper. [1]

Appendix B

Feature variable descriptive statistics, and linear inter-correlations among comment features (N = 2883)

Variable	1	2	3	4	5	6	7	8	9	10	11	12
Mean	3.20	0.30	0.90	0.20	0.80	1.10	0.40	0.60	0.20	0.04	0.60	0.40
Standard deviation	3.20	0.70	1.60	0.50	1.60	2.50	0.80	0.90	0.60	0.20	0.80	0.90
%zero	12%	73%	49%	82%	69%	70%	78%	64%	84%	98%	69%	83%
Maximum	31	7	9	3	8	10	5	6	5	1	4	5

Variable	1	2	3	4	5	6	7	8	9	10	11	12
Feature intercorrelations												
1. Subdimension	–											
2. Open Questions	−0.00	–										
3. Binary Questions	0.46***	0.09***	–									
4. Strengths ident	−0.17***	0.25***	−0.09***	–								
5. Weakness ident	0.40***	−0.04*	−0.16***	0.13***	–							
6. Expectation	0.47***	−0.07***	−0.07***	0.06**	0.88**	–						
7. Quality review	−0.11***	−0.06***	−0.07***	0.51***	0.20***	0.26***	–					
8. Suggestion	0.24***	0.00	−0.12***	0.13***	0.72**	0.70**	0.31***	–				
9. Specific number	−0.00	−0.04	−0.06**	0.07***	−0.02	0.02	0.28***	−0.08***	–			
10. Summary	0.26***	−0.02	0.03	0.09***	0.31***	0.27***	0.07***	0.26***	−0.05**	–		
11. Elaboration	0.22***	−0.06**	−0.12***	0.20***	0.63**	0.57***	0.25***	0.61***	−0.04*	0.21***	–	
12. Example	0.46***	−0.06***	−0.09***	−0.03	0.76**	0.70**	−0.03	0.53***	0.04*	0.38***	0.44***	–

Note. *p < 0.05, **p < 0.01, ***p < 0.001.

Acknowledgements
Thanks to Qiuchen Yu for reliability data coding.

Author contributions
Huifeng Mu: Data curation; Formal analysis; Visualization; Writing—original draft; Writing—review & editing. Christian D. Schunn: Supervision; Conceptualization; Methodology; Resources; Software; Writing—review & editing.

Funding
The research was funded by the China Scholarship Council under grant [202008330496].

Availability of data and materials
The dataset used to support the findings of this study is available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate
Analysis of the data was approved by the University of Pittsburgh Human Research Protection Office.

Competing interests
Christian D. Schunn is a co-inventor of the peer review system used in the study. There are no conflicts of interest, as this study addresses general research questions rather than evaluating a particular product.

Received: 16 February 2024 Accepted: 8 January 2025
Published online: 31 January 2025

References

Alemdag, E., & Yildirim, Z. (2022). Effectiveness of online regulation scaffolds on peer feedback provision and uptake: a mixed methods study. *Computers & Education, 188*, 104574.

Alonso-Tapia, J., & Panadero, E. (2010). Effects of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje, 33*(3), 385–397.

Alqassab, M., Strijbos, J. W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11–30.
Belland, B. R. (2016). Instructional scaffolding in STEM education: strategies and efficacy evidence. *Springer International Publishing*. https://doi.org/10.1007/978-3-319-02565-0
Belland, B. R., Kim, C., & Hannafin, M. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist, 48*(4), 243–270. https://doi.org/10.1080/00461520.2013.838920
Bong, J., & Park, M. S. (2020). Peer assessment of contributions and learning processes in group projects: an analysis of information technology undergraduate students' performance. *Assessment & Evaluation in Higher Education, 45*(8), 1155–1168.
Carson, L., & Kavish, D. (2018). Scaffolding rubrics to improve student writing: preliminary results of using rubrics in a sociology program to enhance learning and mechanical writing skills. *Societies, 8*, 34. https://doi.org/10.3390/soc8020034
Chang, C. Y. H. (2016). Two decades of research in L2 peer review. *Journal of Writing Research, 8*(1), 81–117.
Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328–338.
Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: a web-based reciprocal peer review system. *Computers & Education, 48*(3), 409–426.
Cook, A., Dow, S., & Hammer, J. (2020, July). Designing interactive scaffolds to encourage reflection on peer feedback. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 1143–1153).
Cowie, B. (2005). Pupil commentary on assessment for learning. *Curriculum Journal, 16*(2), 137–151.
Crossley, S. A. (2020). Linguistic features in writing quality and development: an overview. *Journal of Writing Research, 11*(3), 415–443.
Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: use of diagrams in computer-based training environments. *Instructional Science, 30*(6), 433–464. https://doi.org/10.1023/A:1020516301541
Cui, Y., & Schunn, C. D. (2024). Peer feedback that consistently supports learning to write and read: providing comments on meaning-level issues. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2024.2364025
Dawson, P., Yan, Z., Lipnevich, A., Tai, J., Boud, D., & Mahoney, P. (2024). Measuring what learners do in feedback: the feedback literacy behaviour scale. *Assessment & Evaluation in Higher Education, 49*(3), 348–362.
Deci, E. L., & Ryan, R. M. (2012). Motivation, personality, and development within embedded social contexts: An overview of self-determination theory. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 85–107). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0006.
Deiglmayr, A. (2018). Instructional scaffolds for learning from formative peer assessment: effects of core task, peer feedback, and dialogue. *European Journal of Psychology of Education, 33*(1), 185–198.
Dmoshinskaia, N., Gijlers, H., & de Jong, T. (2021). Giving feedback on peers' concept maps in an inquiry learning context: the effect of providing assessment criteria. *Journal of Science Education and Technology, 30*(3), 420–430.
Dong, Z., Gao, Y., & Schunn, C. D. (2023). Assessing students' peer feedback literacy in writing: Scale development and validation. *Assessment & Evaluation in Higher Education, 48*(8), 1103–1118.
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.
Double, K. S., McGrane, J. A., Stiff, J. C., & Hopfenbeck, T. N. (2019). The importance of early phonics improvements for predicting later reading comprehension. *British Educational Research Journal, 45*(6), 1220–1234.
Ellegaard, M., Damsgaard, L., Bruun, J., & Johannsen, B. F. (2018). Patterns in the form of formative feedback and student response. *Assessment & Evaluation in Higher Education, 43*(5), 727–744.
Ellis, R. (2013). Corrective feedback in teacher guides and SLA. *Iranian Journal of Language Teaching Research, 1*, 1–18.
Ertmer, P. A., Richardson, J. C., Lehman, J. D., Newby, T. J., Cheng, X., Mong, C., & Sadaf, A. (2010). Peer feedback in a large undergraduate blended course: perceptions of value and learning. *Journal of Educational Computing Research, 43*(1), 67–88.
Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101*(3), 577.
Fleckenstein, J., Meyer, J., Jansen, T., Keller, S., & Köller, O. (2020). Is a long essay always a good essay? The effect of text length on writing assessment. *Frontiers in Psychology, 11*, 562462.
Gan, M. J., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science, 42*(6), 861–878.
Gao, Y., An, Q., & Schunn, C. D. (2023). The bilateral benefits of providing and receiving peer feedback in academic writing across varying L2 proficiency. *Studies in Educational Evaluation, 77*, 101252.
Ge, X., Chen, C. H., & Davis, K. A. (2005). Scaffolding novice instructional designers' problem-solving processes using question prompts in a web-based learning environment. *Journal of Educational Computing Research, 33*(2), 219–248.
Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*(5), 435–449.
Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about teaching in higher Ed: neglected opportunities to promote change. *CBE Life Sciences Education, 13*(2), 187–199. https://doi.org/10.1187/cbe.13-12-0235
Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104–137.
Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment & Evaluation in Higher Education, 40*(1), 151–164.
Hannafin, M., Land, S., & Oliver, K. (1999). Open-ended learning environments: Foundations, methods, and models. A new paradigm of instructional theoryIn C. M. Reigeluth (Ed.), *Instructional design theories and models* (Vol. II, pp. 115–140). Lawrence Erlbaum Associates.
Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology, 34*(3), 269–290.

Hasan, M., & Rezaul Karim, M. (2019). Scaffolding effects on writing acquisition skills in EFL context. *Arab World English Journal, 10*, 11.
Hill, J., Berlin, K., Choate, J., Cravens-Brown, L., McKendrick-Calder, L., & Smith, S. (2021). Exploring the emotional responses of undergraduate students to assessment feedback: implications for instructors. *Teaching and Learning Inquiry: THe ISSOTL Journal, 9*(1), 294–316.
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.
Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2017). Peer feedback on college students' writing: exploring the relation between students' ability match, feedback quality and essay performance. *Higher Education Research & Development, 36*(7), 1433–1447.
Jiang, X., & Ironsi, S. S. (2024). Do learners learn from corrective peer feedback? Insights from students. *Studies in Educational Evaluation, 83*, 101385. https://doi.org/10.1016/j.stueduc.2024.101385
Jin, X., Jiang, Q., Xiong, W., Feng, Y., & Zhao, W. (2022). Effects of student engagement in peer feedback on writing performance in higher education. *Interactive Learning Environments, 32*, 1–16.
Jurkowski, S. (2018). Do question prompts support students in working with peer feedback? *International Journal of Educational Research, 92*, 1–9.
Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1–19.
Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., Van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: a synthesis of the literature. *Education and Information Technologies, 29*(1), 763–813.
Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review, 30*, 397–429.
King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*(1), 111–126.
King, A. (1994). Guiding knowledge construction in the classroom: effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2), 338–368.
King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice, 41*(1), 33–39.
King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *The Journal of Experimental Education, 61*(2), 127–148.
Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, R. J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning, 13*, 213–233.
Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(2), 211–224.
Könings, K. D., van Zundert, M., & van Merriënboer, J. J. (2019). Scaffolding peer-assessment skills: Risk of interference with learning domain-specific skills? *Learning and Instruction, 60*, 85–94.
Lamb, R. L., & Etopio, E. (2019). Virtual reality simulations and writing: a neuroimaging study in science education. *Journal of Science Education and Technology, 28*(5), 542–552.
Latifi, S., Noroozi, O., & Talaee, E. (2021). Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes. *British Journal of Educational Technology, 52*(2), 768–784.
Lee, J. E., & Recker, M. (2021). The effects of instructors' use of online discussions strategies on student participation and performance in university online introductory mathematics courses. *Computers & Education, 162*, 104084.
Lee, Y. F., Lin, C. J., Hwang, G. J., Fu, Q. K., & Tseng, W. H. (2021). Effects of a mobile-based progressive peer-feedback scaffolding strategy on students' creative thinking performance, metacognitive awareness, and learning attitude. *Interactive Learning Environments, 31*, 1–17.
Leijen, D. A. (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.
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, 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.
Little, T., Dawson, P., Boud, D., & Tai, J. (2024). Can students' feedback literacy be improved? A scoping review of interventions. *Assessment & Evaluation in Higher Education, 49*(1), 39–52.
Liu, Y. (2019). Using reflections and questioning to engage and challenge online graduate learners in education. *Research and Practice in Technology Enhanced Learning, 14*(1), 1–10.
MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing, 32*(6), 1553–1574.
Misiejuk, K., & Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers & Education, 175*, 104319.
Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: a systematic review. *Review of Education*. https://doi.org/10.1002/rev3.3292
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.
Neubaum, G., Wichmann, A., Eimler, S. C., & Krämer, N. C. (2014). Investigating incentives for students to provide peer feedback in a semi-open online course: an experimental study. In *Proceedings of The International Symposium on Open Collaboration* (pp. 1–7).
Nguyen, H., Xiong, W., & Litman, D. (2016). Instant feedback for increasing the presence of solutions in peer reviews. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 6–10).

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122.
Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*(3), 259–271.
Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2014). Rubrics vs. self-assessment scripts: effects on first year university students' self-regulation and performance/Rúbricas y guiones de autoevaluación: efectos sobre la autorregulación y el rendimiento de estudiantes universitarios de primer año. *Infancia y Aprendizaje, 37*(1), 149–183.
Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806–813.
Paris, B. (2022). Instructors' perspectives of challenges and barriers to providing effective feedback. *Teaching and Learning Inquiry*. https://doi.org/10.20343/teachlearninqu.10.3
Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education, 43*(12), 2263–2278.
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.
Peters, O., Körndle, H., & Narciss, S. (2018). Effects of a formative assessment script on how vocational students generate formative feedback to a peer's or their own performance. *European Journal of Psychology of Education, 33*(1), 117–143.
Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: twenty-five years of progress* (pp. 263–305). Lawrence Erlbaum Associates Publishers.
Rietsche, R., Caines, A., Schramm, C., Pfütze, D., & Buttery, P. (2022, July). The specificity and helpfulness of peer-to-peer feedback in higher education. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 107–117).
Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*(1), 75–99.
Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*, 345–372. https://doi.org/10.1002/sce.10130
Schunn, C. (2016). Writing to learn and learning to write through SWoRD. In Adaptive educational technologies for literacy instruction (pp. 243–260). Routledge.
Shiu, A. T., Chan, C. W., Lam, P., Lee, J., & Kwong, A. N. (2012). Baccalaureate nursing students' perceptions of peer assessment of individual contributions to a group project: a case study. *Nurse Education Today, 32*(3), 214–218.
Shvidko, E. (2020). Taking into account interpersonal aspects of teacher feedback: principles of responding to student writing. *Journal on Empowering Teaching Excellence, 4*(2), 7.
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.
Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: a study of high school computer course. *Computers & Education, 49*(4), 1161–1174.
Tuckman, B. W. (2007). The effect of motivational scaffolding on procrastinators' distance learning outcomes. *Computers & Education, 49*(2), 414–422. https://doi.org/10.1016/j.compedu.2005.10.002
Viberg, O., Baars, M., Mello, R. F., Weerheim, N., Spikol, D., Bogdan, C., Gasevic, D., & Paas, F. (2024). Exploring the nature of peer feedback: an epistemic network analysis approach. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.13035
Vuogan, A., & Li, S. (2022). Examining the effectiveness of peer feedback in second language writing: a meta-analysis. *TESOL Quarterly, 57*(4), 1115–1138.
Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education, 33*(1), 165–184.
Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015
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. (2021a). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal, 58*(3), 492–526.
Wu, Y., & Schunn, C. D. (2021b). 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. (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*, 118.
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.
Xun, G. E., & Land, S. M. (2004). A conceptual framework for scaffolding III-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development, 52*(2), 5–22.
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. https://doi.org/10.1016/j.chb.2023.107857
Zhang, Y., Schunn, C. D., & Wu, Y. (2024). What does it mean to be good at peer reviewing? A multidimensional scaling and cluster analysis study of behavioral indicators of peer feedback literacy. *International Journal of Educational Technology in Higher Education, 21*(1), 1–22. https://doi.org/10.1186/s41239-024-00458-1
Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: a meta-analysis. *Asia Pacific Education Review, 17*, 187–202.

Zong, Z., Schunn, C. D., & Wang, Y. (2021a). Learning to improve the quality peer feedback through experience with peer feedback. *Assessment & Evaluation in Higher Education, 46*(6), 973–992.
Zong, Z., Schunn, C. D., & Wang, Y. (2021b). What aspects of online peer feedback robustly predict growth in students' task performance? *Computers in Human Behavior, 124*, 106924.
Zong, Z., Schunn, C., & Wang, Y. (2022). What makes students contribute more peer feedback? The role of within-course experience with peer feedback. *Assessment & Evaluation in Higher Education, 47*(6), 972–983.

Publisher's Note