Does Learning from Giving Feedback Depend on the Product Being Reviewed: Concept Maps or Answers to Test Questions?

Natasha Dmoshinskaia · Hannie Gijlers · Ton de Jong

Accepted: 8 September 2021 / Published online: 30 September 2021
© The Author(s) 2021

Abstract
Giving feedback to peers can be a valuable learning experience for a feedback provider. However, different types of products require different types of feedback, which, in turn, may lead to different learning outcomes. The current study investigates the effect on the learning of feedback providers of reviewing different types of products. Secondary school students (n = 127) were randomly assigned to one of the two conditions: giving peer feedback on either concept maps or answers to open-ended test questions. Both types of product, created by the researchers, were comparable with regard to content: they included the same misconceptions and were both of average quality. Giving peer feedback was part of a chemistry lesson delivered in an online inquiry learning environment. Students’ post-test scores, their own learning products, and the quality of the provided feedback were analysed to check for an effect on learning. There was no difference in post-test scores between the conditions, but the quality of the provided feedback predicted post-test scores. This indicates that it is not the type of product reviewed that matters, but the effort that students put into giving feedback. Possible implications for practice and further research directions are discussed.

Keywords Peer assessment · Peer feedback · Concept maps · Test answers · Inquiry learning

Introduction
Peer assessment has a positive effect on learning, especially if the process is computer-mediated (Li et al., 2020). This became clear from a recent meta-analysis conducted by Zheng et al. (2020) that included 37 empirical studies of technology-facilitated peer assessment. Their findings demonstrated a significant and medium-sized effect of technology-supported peer assessment on learning. Students involved in peer assessment had higher test scores and created higher quality learning products than those who were not involved.

Peer Feedback
Peer feedback is part of peer assessment, inasmuch as peer assessment usually consists of two processes: giving feedback to and receiving feedback from peers. Overall, the former is considered to be more beneficial for learning than the latter (Ion et al., 2019; Li & Grion, 2019; Phillips, 2016). This can be attributed to the fact that giving feedback (that is, reviewing) requires more active cognitive involvement with the product and the material than receiving feedback. When reviewing a peer’s work, students must perform several cognitive activities. Not only do they have to think about the important characteristics of the reviewed product and evaluate whether the product lacks any of those characteristics, but they also need to think of ways to improve the product accordingly. Moreover, if students give feedback on the same type of learning product that they must produce themselves, they get an opportunity to see the same product completed by peers who used a potentially different approach or adopted a different angle, which can enrich their understanding of the topic. For example, Wu and Schunn (2020) found that having secondary school students provide
comments on writing drafts led to their revising of their own drafts and to more learning.

A number of studies have shown that the quality of the product affects how much a feedback provider (that is, a reviewer) learns (Alqassab et al., 2018; Patchan & Schunn, 2015; Tsivitanidou et al., 2018). One explanation for such an influence is that the quality of the product determines the type and amount of feedback. For instance, feedback on a lower quality product would involve identifying more problems and suggesting more solutions than giving feedback on a higher quality product, while spotting one mistake in a product of generally good quality could be more challenging than finding many mistakes in a lower quality product. There can also be an interaction effect between the reviewed product’s level of quality and the reviewer’s prior knowledge. Because students with higher prior knowledge understand the topic and potentially know the product better, they might be better able to find mistakes and suggest solutions for improving reviewed products than students with lower prior knowledge. These considerations are important, as the quality of the feedback they give influences reviewers’ learning, which was shown, for example, in a study by Li et al. (2010) demonstrating that the quality of the provided feedback had a significant relationship with reviewers’ own learning products.

The influence of the reviewed product’s characteristics on the reviewer’s learning might also be true for the type of product reviewed, as processing information presented in different forms may require different cognitive activities (e.g., Kalyuga & Plass, 2017), so the feedback can be different. One type of product can stimulate more higher-order thinking than another type, and thereby lead to different learning gains for the reviewer. Students’ understanding of a topic can be demonstrated in different types of products, but these products can present that understanding in different ways. Comparing the process of giving feedback and the feedback given for different products can shed some light on the learning that originates from giving feedback, which will be explored in the next section.

**Type of Products Reviewed**

The current study aims to investigate the role of the type of reviewed product in the reviewer’s learning by comparing the process and results of giving feedback for two types of products: concept maps and answers to open-ended test questions. Both products were of a smaller scale (compared to papers or essays), reflected understanding of a topic and were used in secondary education. Concept mapping is a way of presenting a domain topic through concepts and the connections between them. Concept maps contribute to deeper understanding of a topic, as they encourage students to think about such connections (e.g., Novak, 2010). Schroeder et al. (2018) conducted a meta-analysis on the effect of concept mapping on learning, which included 63 studies. The authors demonstrated a moderate overall effect of concept mapping on learning; moreover, this effect was found across different domains. The same meta-analysis found that working with concept maps was especially beneficial for secondary school students.

As concept maps provide a way to visualise knowledge through relations between key concepts, which was shown to be good for learning, concept maps should be a good product to review in order to learn from that process. Indeed, for example, Chen and Allen (2017) demonstrated that not only did concept mapping itself lead to more conceptual understanding, but also students who reviewed peers’ concept maps learned more about the topic than students who did not participate in reviewing. However, reviewing peers’ concept maps can be a rather challenging task for students. Concept mapping itself is a skill that needs to be developed, and giving feedback on such a product requires deeper understanding of the topic (e.g., Cañas et al., 2017). The general and aggregated view presented by concept maps can also be an obstacle for some students when trying to provide feedback, as reviewing a concept map could require greater competence from them: their prior knowledge should be complete enough to understand the connections or provide suggestions for improvement (e.g., Novak & Cañas, 2006). This means that reviewing concept maps may not be equally beneficial for students with different levels of prior knowledge.

The other type of reviewed product was answers to open-ended test questions, as the questions may be similar to concept maps in terms of presenting the main concepts related to the topic and connections between them. This means that such answers may require deeper thinking, as do concept maps. However, the difference is in the way information is presented: answers to open-ended test questions present information more explicitly than concept maps, which may make a mistake more obvious and the process of giving feedback easier. Giving feedback on answers to open-ended test questions includes identifying mistakes in the answers, which is also called an error-detection task (Adams et al., 2019). Such tasks are widely used in many school subjects: students need to find a mistake in a given sentence or example, explain what the mistake is and suggest a correction. This makes answers to test questions much more common and familiar products to review. Moreover, according to Adams et al. (2019), training students in error-detection tasks leads not only to their identifying more mistakes, but also to their giving better feedback to peers about those mistakes. In other words, if students practice these tasks often enough, they can give better feedback. To sum up, answers to test questions may present information in a more straightforward way and be more familiar for students than concept maps, which may make them an easier product.
to review. However, reviewing answers to open-ended test questions might not create a cognitive challenge equal to that of reviewing concept maps. Moreover, the experience of such challenge can depend on students’ prior knowledge. Altogether, this may mean that learning that arises from giving feedback on concept maps and on answers to open-ended test questions can be different.

Research Questions

As described above, giving feedback on concept maps can stimulate more and deeper thinking than reviewing answers to open-ended test questions, but concept maps can also be more difficult to review. Answers to open-ended test questions present information in a more common and straightforward way that is easier to review, but, as a consequence, they may require less in the way of deep thinking.

This leads to the following research questions: How does the type of reviewed product (concept maps or answers to open-ended test questions) affect peer reviewers’ learning? Is there a differential effect for students with different levels of prior knowledge?

As learning can be measured through different products, in the current study, learning is assessed through several outcomes of students’ work: post-test scores, the quality of students’ own concept maps, and the quality of the provided feedback.

Method

Participants

The sample consisted of 157 Dutch secondary school students (third grade in the Dutch school system) from the same school, with an average age of 14.89 years ($SD = 0.40$). Students in six classes took part in the experiment. The criteria for inclusion, apart from being a student in a participating class, were attending both lessons that were part of the experiment, completing the pre-test and post-test, and providing feedback. Applying these criteria reduced the group to 127 students (59 girls and 68 boys). The majority of excluded students missed one or both lessons because of illness.

In each class, students were randomly assigned to one of the two conditions: giving feedback on concept maps (CM condition), 66 students; and giving feedback on answers to open-ended test questions (test condition), 61 students.

Study Design

In the current study, participants were asked to give feedback on (fictitious) peers’ learning products. In both conditions, they were supported in this process by assessment criteria that were relevant to the product type. Such criteria aimed at making the process of giving feedback more natural than when standard assessment criteria that relate poorly to the particular product are used. In the CM condition, students were asked to give feedback on two concept maps using assessment criteria presented in a question form, which guided students in the direction of the desired features of a concept map. The assessment criteria were rather general and not specific for the domain. This approach was based on the study by van Dijk and Lazonder (2013), in which the most important characteristics of a concept map were discussed. The questions used in this study as guidelines for giving feedback on concept maps are presented in Table 1.

In the test condition, students were asked to give feedback on answers to open-ended test questions and guided through this process, as they needed to indicate if the answer was correct or not and, in case of an incorrect answer, to explain what the mistake was. These answers to test questions covered the same concepts, relations, and misconceptions as the two concept maps in the other condition.

Procedure and Materials

The lessons were given in the students’ native language (Dutch) and followed the national curriculum for chemistry. The topic fit into the regular programme of study, and presented the theme of matter and elements, and in particular, atomic structure. The experiment included two sessions (50 min each) and took place during two successive chemistry lessons, according to the regular timetable of the participating classes. These sessions occurred within 5 days for all classes. All tests (pre- and post-) and working in the ILS (see below) were done online, with each participant working individually on a computer.

At the beginning of the experiment, a brief introduction to the goals, procedure, and privacy rules was given. Students could ask to withdraw their data from the analysis, but they still had to complete the tasks as part of their learning programme. The researcher was present during the lessons; participants could ask questions about the procedure or tools, but not about the content.

| Table 1 | Assessment criteria for giving feedback on concept maps (translated from Dutch) |
|---------|--------------------------------------------------------------------------------|
| 1.      | What important concepts are missing?                                           |
| 2.      | How would you change the structure of the map?                                |
| 3.      | Which links should be renamed to be more meaningful?                          |
| 4.      | What examples should be added?                                                |
| 5.      | Why is this concept map helpful or not helpful for understanding the topic?     |
The lessons were built with the help of the Go-Lab ecosystem (see www.golabz.eu). This ecosystem allows the creation of inquiry learning spaces (ILSs) that guide students through the inquiry process (de Jong et al., 2021). By following an inquiry learning cycle, students explore a scientific phenomenon in a way resembling scientific research. The cycle includes several phases (orientation, conceptualization, investigation, conclusion, and discussion), each having a particular purpose in the inquiry process (Pedaste et al., 2015).

The timeline of the experiment and the content of each lesson are shown in Fig. 1.

The Concept Mapper tool and the online lab used during the lessons are shown in Figs. 2 and 3.

All reviewed learning products (concept maps and test answers) were created by the research team. This was done for experimental control to ensure that participants assessed the same products and had the same opportunities for learning from giving feedback on them, as the quality of the reviewed product influences the feedback-giving process and its outcomes (see, e.g., Patchan & Schunn, 2015). To create a more realistic setting, students were told that the products came from other students, probably from a different school. The products for both conditions were of medium quality, presenting both right and wrong information. Moreover, products for both conditions included the same three misconceptions: an electron has a mass, a neutron has a charge, and neutrons define the element.

In the CM condition, students gave feedback on two concept maps, one after the other. We decided to split one complex concept map into two to make them more manageable for students. Apart from misconceptions, both concept maps had some important concepts missing, and not all of the concepts were connected; in other words, they both had some room for improvement.

In the test condition, participants commented on answers to test questions. The answers to the five open-ended questions included two correct and three incorrect answers. Students first had to identify the incorrect ones and then had to explain what was wrong with the answer for those cases.

Students gave their feedback anonymously using a special peer assessment tool. The tool allowed them to see the reviewed products and the assessment criteria and to write their comments. The view of the peer assessment tool for both conditions is presented in Fig. 4a, b.

---

**Fig. 1** Timeline and content of the experimental sessions (the names are translated from Dutch)
Pre-tests and post-tests were parallel versions and consisted of six open-ended questions each, with a maximum score of 11 points. The tests were considered parallel, as they addressed the same concepts but in differently formulated questions. Cronbach’s alpha was 0.43 for the pre-test and 0.61 for the post-test. The low value for the pre-test could be explained by the very low pre-test scores (see Table 4). The value for the post-test was also on the lower side, which can be connected to the fact that the scores were not that high (on average 6 out of 11). Tests were developed to check knowledge and understanding of the topic of the ILS only, and did not include potential extra items checking the same construct. There were three types of questions used: those checking knowledge (1 point), those requiring application of knowledge (2 points), and those requiring understanding the connections between the concepts (3 points). Examples of each type of question are given in Table 2. Moreover, for questions with more than one point, a coding scheme was used for a partially correct answer. It awarded more points for argumentation explaining the connection between the concepts (see Table 3 for an example).

Analysis

First, the pre-test and post-test consisted of open-ended questions that needed to be scored. The scoring was done by applying the coding scheme, which was developed by the researchers and approved by participating teachers. The scheme included the right answers and corresponding points. A second rater graded 20% of the pre-tests and post-tests to estimate the inter-rater reliability, which was found to be good, with Cohen’s kappa being 0.92 for the pre-test and 0.74 for the post-test.

Second, the final versions of the students’ own concept maps (after giving feedback) were coded. The fact of students’ changing their concept maps after giving feedback was coded as 1, while not changing was coded as 0. The following criteria were used to code characteristics of the concept maps:

- Proposition accuracy score — the number of correct links;
- Salience score — the proportion of correct links out of all links used;
- Complexity score — the level of complexity.

The coding aimed at evaluating the quality of the concept maps through the main concepts and links between them, and not through comparing the student’s concept map with an exert concept map. This approach was chosen because students could have very different ways of presenting their ideas in a concept map, and any way was considered potentially valid.

The type of scale used and maximum number of points differed per score. The proposition accuracy score had no set maximum, as students could not only use pre-defined concepts, but also add their own relevant concepts and create new links. The maximum for the salience score was one, by definition. Finally, the complexity score aimed to discriminate concept maps with different structures. A linear construction (“sun” or “snake” shaped) would get one point, while a hierarchical structure with more than one level would get two points.

A second rater coded 20% of the concept maps, with Cohen’s kappa reaching 0.62. This is an acceptable yet moderate result. As the scales for assessed characteristics were continuous scales and Cohen’s kappa is less informative for a continuous scale variable, a Pearson’s correlation was also used to check the inter-rater agreement, with $r = 0.95$ ($p < 0.01$). Together with the Cohen’s kappa value, this suggested that the scoring was reliable.

Third, the quality of the feedback given by students was evaluated. For the concept maps condition, the aim of the feedback was to provide correct and/or (potentially) useful comments characterising the concept map that was reviewed. A comment was coded as correct if it was accurate in terms of the domain (chemistry) and as useful if it also contained an explanation from the domain perspective. Each correct and useful answer received one point; any additional suggestion for the same question received half a point. Feedback for each concept map was assessed...
separately, and the average of the two scores was used as the final score for the quality of feedback. Taking into account the main missing concepts and mistakes included in the concept maps, the maximum score for the quality of the feedback was expected to be eight points.

For the test condition, the aim of the feedback was to identify incorrect answers and provide an explanation of what was wrong. One point was given for each accurate identification of the correctness/incorrectness of an answer and one point for each viable explanation of why an answer was incorrect. With three incorrect

Fig. 3 View of the online lab (English version, same as the Dutch version used in the lesson). Images by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY 4.0 https://phet.colorado.edu

Fig. 4  a View of the peer assessment tool for the CM condition (translated from Dutch).  b View of the peer assessment tool for the test condition (translated from Dutch)
answers out of five, the maximum score for the quality of feedback was eight points. A second rater coded 20% of the feedback given for both conditions, with Cohen’s kappa being 0.86.

**Results**

The difference in pre-test scores between the conditions was not significant: $M_{CM} = 2.00$ ($SD = 1.47$) and $M_{TEST} = 1.97$ ($SD = 1.73$), $\kappa(125) = 0.12$, $p = .91$ (see Table 4).

For further analysis, participants were divided into three groups based on their pre-test results. The groups were low prior knowledge (pre-test score lower than 1 SD below the mean for the entire sample), average prior knowledge (pre-test score within 1 SD above or below the mean), and high prior knowledge (pre-test score higher than 1 SD above the mean). The overall distribution of students among the low, average, and high prior knowledge groups in our sample was 14, 90, and 23 students, respectively.

Descriptive statistics for all prior knowledge groups and conditions are presented in Table 4.

**Learning Gain**

As a prerequisite for further analyses, whether students learned during the experimental lessons was first checked. As post-test scores were not normally distributed, a non-parametric Wilcoxon signed-rank test was used for pre-test and post-test scores. The results showed that students learned during the lessons, as a whole group and in each condition ($Z_{CM} = 6.80$, $p < .01$; $Z_{TEST} = 6.53$, $p < .01$; $Z = 9.41$, $p < .01$).

**Effect of Condition on Post-test Scores**

Second, an ANOVA was conducted to answer the research questions about the influence of the type of product reviewed (concept maps or answers to test questions) and prior knowledge on the post-test scores. Condition and prior knowledge level were used as independent variables and post-test score as the dependent variable. No main effect was found to be statistically significant for either of the variables, nor was there an interaction effect. The pairwise comparisons were not significant either.

**The Quality of Students’ Own Concept Maps**

As students in one condition reviewed concept maps and students in the other condition reviewed answers to test questions, but students in both conditions had to create their own concept maps, we expected there to be a difference between the conditions in the quality of the final concept maps produced, which was checked with a non-parametric independent-samples Mann–Whitney $U$-test. The descriptive statistics for the variables used are presented in Table 5. No statistically significant differences were found for any of these variables.

A regression analysis was conducted to see if the quality of what students produced while working in the ILS — their own concept maps and feedback given on fictitious peers’ products – predicted their post-test scores. Apart from the characteristics of the students’ concept maps, the fact of changing their concept map after giving feedback was also included in the analysis, as well as the quality of the feedback given and students’ prior knowledge. The coefficients for the independent variables included in the analysis are shown in Table 6. As for the quality of students’ own concept maps, proposition accuracy (number of correct links) was found to be a significant predictor of post-test scores, with the post-test score increasing by 0.35 points when the proposition accuracy increased by 1 point and all other variables stayed the same.

**The Quality of Feedback Given**

Even though the products to give feedback on were different in the two conditions, students in both conditions were supported in the process of giving feedback by assessment criteria matching the product type. A non-parametric independent-samples Mann–Whitney $U$-test showed that the difference in feedback quality was significant ($U = 1547.50$, $p = .046$), with feedback on answers to test questions being of higher quality overall than feedback on concept maps.\(^1\)

\(^1\) Due to technical reasons, the feedback from two participants was not saved, so the feedback of 125 participants was used for this analysis.
As the regression analysis above demonstrated (see Table 6), the quality of the feedback given was a significant predictor of post-test scores, with the post-test score increasing by 0.47 points when the quality of feedback increased by 1 point and all other variables stayed the same.

The statistically significant predictors of the post-test score (i.e., proposition accuracy score and quality of feedback given) were not found to be correlated with prior knowledge ($p_{\text{ACC}} = .064$, $p_{\text{FEEDBACK}} = .158$).

### Conclusion and Discussion

The current study aimed to determine whether reviewing different types of products (in our case, concept maps and answers to open-ended test questions) would lead to different learning results. Below, we present conclusions based on the results obtained and their interpretation for practice.

First, it was expected that post-test scores might differ between the conditions, as the reviewed products (concept maps and answers to test questions) differ in the way they present information and in their level of familiarity for students. Our review of literature did not give a clear indication of which condition should show higher results. On the one hand, reviewing concept maps might lead to better conceptual learning (and thus, higher post-test scores) than reviewing answers to test questions, as concept maps help to visualise relationships between key concepts (e.g., Chen & Allen, 2017; Schroeder et al., 2018). On the other hand, test answers are more familiar to students and present information in a more straightforward way than different learning results. Below, we present conclusions based on the results obtained and their interpretation for practice.
concept maps, which might lead to noticing and explaining more mistakes and thus to better learning (e.g., Adams et al., 2019). However, the conditions did not show a statistically significant difference in post-test scores (with differences in pre-test scores not being statistically significant either). Finding no difference may indicate that reviewing both types of learning products can be beneficial for reviewers’ learning. This can make implementation of peer feedback in the classroom easier to do and wider in its application.

A surprising finding was that prior knowledge level did not explain post-test scores. A possible explanation is that the pre-test results were so low that they did not matter that much for the majority of the population (the low and average prior knowledge groups). For the low prior knowledge group, the average pre-test score was 0.00, which might be attributed to the specificity of the topic — atomic structure. Even though students had studied some material about molecules and atoms before, they had never studied this particular topic or the terms associated with it. One cannot answer a question about the influence of adding a proton to an atom if the term ‘proton’ is not known. However, the same question makes much more sense if the terms are familiar. This is indirectly supported by the fact that students in both conditions and in all prior knowledge groups did learn during the experiment. Moreover, their learning followed a normal trend, with the average post-test score reaching around 60% of the maximum score.

Second, a difference in the quality of students’ own concept maps by condition was expected. Students in both conditions had to create a concept map before they worked in the online lab and could rework it after giving feedback on fictitious peers’ products. However, students from one condition could rework their own concept maps after reviewing fictitious peers’ concept maps, whereas students from the other condition could do so after reviewing answers to test questions. Several studies have shown that reviewing the same type of product that students have to create themselves may lead to revising their own products and thereby to higher quality of their own products (e.g., Li & Grion, 2019; Wu & Schunn, 2020). Therefore, students who reviewed concept maps were expected to have higher quality final concept maps of their own than students who gave feedback on answers to test questions. However, no statistically significant difference was found in the quality of students’ own concept maps between conditions. This might be caused by the fact that both reviewed products (concept maps and answers to test questions) were relatively small and required less time for reviewing compared to products such reports or essays. The amount of time spent interacting with smaller scale products might have not been enough to lead to differences in the quality of reviewers’ own concept maps.

When analysing the characteristics of students’ concept maps, proposition accuracy (the number of correct links) was found to predict post-test scores. This result was quite expected, as the number of correct links showed students’ grasp of the topic when creating concept maps. The same understanding was checked with the post-test. In other words, this result showed that students’ concept maps reflected their current level of knowledge. This is in line with previous studies on concept mapping (Novak & Cañas, 2006; Schroeder et al., 2018). For practice, this can be another supporting argument in favour of using concept maps in schools, as they can demonstrate students’ knowledge and encourage them to think about important connections between the key concepts.

Third, the results showed a significant difference in the quality of the provided feedback between the conditions: students gave better feedback on the answers to test questions than on the concept maps. There are several possible explanations for that. First, the answers to test questions were a much more familiar learning product, so it might have been easier for students to see the mistakes. This is supported by the fact that only 22% of students giving feedback on concept maps spotted at least one misconception placed in the concept maps, while 72% of students giving feedback on the answers to test questions identified at least one mistake, even though these misconceptions and mistakes covered the same content. Second, the lower results for the condition with concept maps could have originated from the complexity of the product. In other words, assessing a concept map included many more aspects, such as structure and missing concepts, while commenting on answers to test questions focused mainly on the correctness of the answer and the reasoning behind it. Moreover, if students found it difficult to create a concept map themselves, they might not feel able to give feedback on peers’ concept maps. Third, even though the coding schemes for both reviewed products awarded points for simple feedback, the assessment criteria for giving feedback on a concept map could require broader understanding than the more straightforward assessment criteria for giving feedback on test answers. Overall, it seemed that spotting mistakes and missing elements in a more complex product such as a concept map was more difficult for students, which led to lower feedback quality. For practice, this can mean that giving students more familiar products to review leads to higher quality feedback, which, in turn, could positively influence learning.

Finally, and probably most importantly, the quality of the feedback provided was found to predict post-test scores: the higher the quality of the provided feedback, the higher the post-test score. An indication that feedback quality could predict post-test scores was also found in one of our previous studies (Dmoshinskaia et al., in press). At that point, there was not enough evidence to say that better feedback led to higher post-test scores.
scores, as they both could have been caused by higher prior knowledge. However, in the current study, prior knowledge was not found to be a significant predictor of post-test scores. Moreover, the quality of the feedback given was not found to be correlated with prior knowledge. This may mean that the quality of feedback indeed explained the post-test score, so that if students invested more effort in giving better feedback, it could lead to more learning for reviewers themselves. This was even more the case for students in the concept map condition, in which giving good feedback was probably more challenging than in the other condition, thus requiring more cognitive involvement. It might seem contradictory to find significant differences in the quality of feedback between the conditions (and the fact that the quality of feedback predicted post-test scores) and not to find a significant difference in post-test scores. A possible explanation is that the variability in post-test scores was too high to reveal an effect of condition. Moreover, quality of feedback explained about 12% of the variability in the post-test scores, indicating the presence of other explanatory factors. The interpretation of the correlation between the quality of provided feedback and the post-test scores for practice may mean that the effect from giving feedback does not depend that much on the product type, as long as students provide meaningful comments. However, this finding should be interpreted with caution as some student-level characteristics (e.g., students’ motivation, engagement with the particular subject, self-regulation skills) could have influenced both the quality of the feedback given and the learning that originated from it. Therefore, more research on giving feedback to peers should be done, with one direction being to study the role of students’ personal characteristics in this process.

The finding that feedback quality can play an important role in learning has also been supported by other studies. For example, the study by Li et al. (2010) demonstrated a significant relationship between the quality of provided feedback and the quality of students’ own projects. The authors suggested that as the ability to give meaningful and constructive feedback is very important for a reviewer, time and effort should be invested in developing this ability. Therefore, an interesting direction for further research would be studying what factors influence such an ability and identifying ways to increase the quality of peer feedback. Another direction could be conducting the same study with different learning products created by students to see if any particular types can stimulate more learning when being reviewed. The current study was the first exploration of the influence that the type of reviewed product can have on learning, so we used two types of products that are frequently used in secondary school education: answers to test questions and concept maps. Investigating giving feedback on other product types may help to better understand this process. Finally, as a feedback-giving activity was embedded in an ILS, it could interesting to study if any particular combination of learning activities during the lesson can provide the most meaningful context for giving feedback.

There are some limitations of the current study. First, the products for reviewing were created by the research team, which means that transfer of the obtained results to a situation with real students’ learning products should be done cautiously, as learning opportunities can differ for products of different types and quality. Second, the moment of giving feedback was quite brief, which could have made the effects less obvious. A longer time of giving feedback or multiple moments of giving feedback could lead to different findings; such a setting could be one direction for future research.

Another aspect worth mentioning is that even though students who gave feedback on answers to test questions provided feedback of higher quality, their learning was not statistically significantly higher than in the other condition. This can be related to small effect sizes for each step of these relationships: between students’ knowledge and the quality of feedback, and between the quality of feedback and post-test scores. A study with more participants could make such relationships more obvious.

In conclusion, several take-away messages emerged from this study. First, the type of reviewed product does not seem to play an important role in reviewers’ learning as long as they provide good quality feedback. Second, even with a rather brief intervention (caused by the use of small-scale products), the quality of peer feedback explained post-test scores, which demonstrates the value of feedback-providing activities for learning. Further research studying peer feedback will be beneficial for educational practice, as knowing more about the conditions under which giving feedback can enhance learning may increase effective use of peer feedback in education.

Acknowledgements We would like to acknowledge the Next-Lab project team members — Jakob Sikken and Tasos Hovardas — for the development of the peer-assessment tool, and Emily Fox for inspiring comments. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

Funding This work was partially funded by the European Union in the context of the Next-Lab innovation action (Grant Agreement 731685) under the Industrial Leadership — Leadership in enabling and industrial technologies — Information and Communication Technologies (ICT) theme of the H2020 Framework Programme.

Declarations

Ethics Approval The approval of the Ethical Committee of the Faculty of Behavioral, Management, and Social Sciences (BMS) University of Twente was obtained prior to the data collection.
Consent to Participate  An appropriate consent form was used for the participants.

Conflict of Interest  The authors declare no competing interests.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Authors Details Omitted for Double-blind Reviewing

Adams, R., Jackson, P., Lenton, K., Dugdale, M., Whittaker, C., Lasry, N., & Charles, E. S. (2019). Error-detection tasks and peer feedback for engaging physics students. Paper presented at the Fifteenth Conference on Education and Training in Optics and Photonics: ETOP 2019, Quebec City, Quebec.

Alqassab, M., Strijbos, J.-W., & Uler, S. (2018). The impact of peer solution quality on peer-feedback provision on geometry proofs: Evidence from eye-movement analysis. Learning and Instruction, 58, 182–192. https://doi.org/10.1016/j.learninstruc.2018.07.003

Cañas, A. J., Reiska, P., & Möllis, A. (2017). Developing higher-order thinking skills with concept mapping: A case of pedagogic frailty. Knowledge Management & E-Learning, 9, 348–365. https://doi.org/10.34105/j.kmel.2017.09.021

Chen, W., & Allen, C. (2017). Concept mapping: Providing assessment of, for, and as learning. Medical Science Educator, 27, 149–153. https://doi.org/10.1007/s40670-016-0365-1

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., & Zacharia, Z. C. (2021). Understanding teacher design practices for digital inquiry-based science learning: The case of Go-Lab. Educational Technology Research & Development. https://doi.org/10.1007/s11423-020-09904-z

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (in press). Giving feedback on peers’ concept maps as a learning experience: Does quality of reviewed concept maps matter? Learning Environments Research. https://doi.org/10.1007/s10908-021-09389-4

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, 124–138. https://doi.org/10.1080/02602938.2018.1484881

Kalyuga, S., & Plass, J. L. (2017). Cognitive load as a local characteristic of cognitive processes. In R. Z. Zheng (Ed.), Cognitive load measurement and application: A theoretical framework for meaningful research and practice (pp. 73–88). Routledge.

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, 193–211. https://doi.org/10.1080/02602938.2019.1620679

Li, L., & Grion, V. (2019). The power of giving feedback and receiving feedback in peer assessment. AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education, 11.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor orassessee: How student learning improves by giving and receiving peer feedback. British Journal of Educational Technology, 41, 525–536. https://doi.org/10.1111/j.1467-8535.2009.00968.x

Novak, J. D. (2010). Learning, creating, and using knowledge: concept maps as facilitative tools in schools and corporations. Routledge.

Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them (Technical report IHMC CmapTools 2006-1). Pensacola, FL: Florida Institute for Human and Machine Cognition.

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, 591–614. https://doi.org/10.1007/s11251-015-9353-x

Pedaste, M., Máeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. Educational Research Review, 14, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Phillips, F. (2016). The power of giving feedback: Outcomes from implementing an online peer assessment system. Issues in Accounting Education, 31, 1–15. https://doi.org/10.2308/iae-50754

Schoeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. Educational Psychology Review, 30, 431–455. https://doi.org/10.1007/s10648-017-9403-9

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. European Journal of Psychology of Education, 33, 51–73. https://doi.org/10.1007/s11251-017-0341-1

van Dijk, A. M., & Lazonder, A. W. (2013). Scaffolding students’ use of learner-generated content in a technology-enhanced inquiry learning environment. Interactive Learning Environments, 21, 194–204. https://doi.org/10.1080/10494820.2013.834828

Wu, Y., & Schunn, C. D. (2020) The effects of providing and receiving peer feedback on writing performance and learning of second-language learners. Journal of Educational Psychology, 112, 710–729. https://doi.org/10.1037/edu0000611

Publisher’s Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.