



# Giving feedback on peers' concept maps as a learning experience: does quality of reviewed concept maps matter?

Natasha Dmoshinskaia<sup>1</sup> · Hannie Gijlers<sup>1</sup> · Ton de Jong<sup>1</sup>

Received: 1 July 2020 / Accepted: 19 September 2021 / Published online: 27 September 2021  
© The Author(s) 2021

## Abstract

Giving feedback to peers can be a powerful learning tool because of the feedback provider's active cognitive involvement with the products to be reviewed. The quality of peers' products is naturally an important factor that might influence not only the quality of the feedback that is given, but also the learning arising from this process. This experimental study investigated the effect of the level of quality of the reviewed product on the knowledge acquisition of feedback providers, as well as the role of prior knowledge in this. Dutch secondary-school students ( $n = 77$ ) were assigned to one of three conditions, which varied in the quality of the learning products (concept maps) on which students had to give feedback while working in an online physics inquiry learning environment. Post-test knowledge scores, the quality of students' own concept maps and the quality of the feedback given were analyzed to determine any effect of condition on the learning of feedback providers. Students providing feedback on the lower-quality concept maps gave better feedback and had higher post-test scores. There was no interaction with level of prior knowledge. Possible implications for practice and further research directions are discussed.

**Keywords** Concept maps · Giving feedback · Inquiry learning · Peer assessment · Quality of reviewed product

## Introduction

Peer assessment should be seen not only as a form of assessment, but also a powerful learning tool. Giving feedback to peers and receiving feedback from them makes students reflect on important elements and characteristics of the products reviewed, thus leading to deeper thinking about the topic. Students who review peers' work and get feedback from peers on their own work create better learning products (e.g. Brill, 2016; Mulder et al., 2014). For these reasons, peer assessment is gaining popularity among teachers, especially when used in online learning environments. A systematic literature review by Tenório et al. (2016) that focused on the phenomenon of peer assessment in online learning showed its

---

✉ Natasha Dmoshinskaia  
n.dmoshinskaia@utwente.nl

<sup>1</sup> Department of Instructional Technology, Faculty of Behavioral, Management, and Social Sciences (BMS), University of Twente, Enschede, The Netherlands

effectiveness at various levels of education. The studies included in the review reported a positive effect of peer assessment on the development of students' skills or their knowledge. Moreover, according to a recent meta-analysis, computer-mediated peer assessment leads to more learning than paper-based peer assessment (Li et al., 2020). This is not surprising, because more can be done in an online learning environment to facilitate organizing the feedback-giving process than in a real classroom. With adaptive tool configuration, students can get more personalized support than with paper-based peer assessment: they can be matched using different strategies (e.g. prior knowledge, skills, product types) even within one class; they can give feedback anonymously or not, and at their own pace. The combination of peer assessment and online learning environments can have advantages not only for implementing peer assessment, but also for working with online environments. Students usually work individually in such environments, and an activity involving giving feedback to peers can make the learning experience richer, because students can see potentially different approaches to dealing with the task.

## Literature review

Providing feedback to peers is normally one part of a peer-assessment process that consists of giving feedback to and receiving feedback from peers. Because giving feedback is an active process, it requires students' active cognitive involvement which, in turn, can increase their feeling of being responsible for and engaged in learning. For example, Ion et al. (2019) found that students perceived giving feedback as a learning experience more than receiving feedback, and that giving feedback helped them to improve their performance. These results are not surprising, because giving feedback includes analyzing the product under review and applying one's existing knowledge in order to identify flaws and suggest solutions. Moreover, giving feedback on peers' work can lead to reflection on one's own work and thus to more learning.

The current study aimed to investigate the learning effect of providing feedback, while taking into consideration two potentially influential factors. The first factor is the quality of the product that is being reviewed. Here, the results of several studies have indicated that students gave better feedback and demonstrated better topic understanding when reviewing a higher-quality product than when reviewing a lower-quality product (e.g. Alqassab et al., 2018a; Tsivitanidou et al., 2018). This means that feedback providers (reviewers) can learn by reviewing high-quality products. Such learning can be attributed to the combination of being introduced to good examples or successful strategies for completing the task and reflecting on them. However, it has also been shown that, in order to evaluate learning products, students must be competent enough to decide what is right and what is wrong (e.g. Alqassab et al., 2018b). Therefore, it could be easier to give feedback on lower-quality product than on higher-quality products, which could lead to higher-quality feedback being given and thus more learning. The cognitive processes triggered by seeing good examples in higher-quality products and by spotting mistakes in lower-quality products could be different, as well as the learning outcomes.

Because products with different levels of quality require different types of comments and suggestions, they can require different levels of domain expertise from a reviewer. This brings us to the second factor that plays a role in the reviewers' learning process: their prior knowledge or ability. Several studies have shown that the levels of reviewed written pieces and reviewers' writing ability influenced the content of their feedback (e.g. Cho & Cho, 2011; Patchan & Schunn, 2015). Their findings indicate that initial ability can matter

in the process of giving feedback. Specifically, the higher it is, the more problems can be identified, whereas strong points are equally easy to spot by reviewers of different ability levels. Thus, there can be an interaction between the reviewer's ability and the quality of the reviewed product. Similar to reviewers' writing ability, their prior knowledge can influence their learning by giving feedback, especially with tasks requiring knowledge rather than skills (e.g. in science education). In terms of giving feedback, prior knowledge can influence the content of the provided feedback; for example, reviewers with low prior knowledge might not be able to understand the written product or come up with useful solutions (Alqassab, 2017). Such a difference in reviewers' performance can lead to a difference in their learning. In the study by Patchan et al. (2013), university students who had to review each other's work were put into two types of pairs according to their domain knowledge levels: either both students were at the same knowledge level; or the two students had different knowledge levels. The findings of this study showed that low-knowledge and high-knowledge reviewers gave similar feedback on a good product (created by a high-knowledge student). However, the feedback on a weak product (created by a low-knowledge student) was different, with more identified problems coming from high-knowledge reviewers. Being able to provide constructive criticism and solutions for the problems identified is connected with the learning process, which can be seen in reviewers' own learning performance. Yalch et al. (2019) demonstrated that the more critical feedback was given by students, the higher grades they received for their own written products. This suggests that students can learn differently from reviewing products of different quality and that students' prior knowledge plays a role in this.

This can be especially important in the context of an online inquiry learning environment, as in the current study. Inquiry learning encourages students to explore a scientific topic and draw their own conclusions about the phenomenon being studied based on the results of experiments that they have performed. Inquiry learning imitates the scientific research process, so that including a feedback-giving activity, which is usually part of that process, can make the experience more natural. A recent study by Dmoshinskaia et al. (2021) demonstrated that commenting on peers' work in an online inquiry learning environment led to reviewers' learning. Online learning environments allow students to follow their own pace and construct their own learning path, and their learning can be enhanced by configuring online tools differently for different groups of students depending on their needs (e.g. de Jong et al. 2021). Therefore, investigating how reviewers with different levels of prior knowledge who are reviewing products of different levels of quality learn from giving feedback can help with formulating practical recommendations for teachers.

## Current study

The present study focused on the roles that the quality of reviewed products and reviewers' prior knowledge play in reviewers' learning. As the previously cited studies show, the quality of the product to be reviewed can influence the quality and content of the feedback provided. Moreover, because the cognitive processes for giving different types of feedback could differ, the quality of the product to be reviewed can influence the learning of a feedback provider. The current study investigated this issue in a secondary-school context. Most of the cited studies focused on feedback given by university students on relatively-large written products—essays and papers. It is not clear as yet if the trends are the same for a smaller type of learning product, such as concept maps, and a younger target population, such as secondary-school children, who were chosen as a target group for the

following reasons. Secondary-school students can give meaningful feedback (Tsivitanidou et al., 2011) and giving feedback to peers can be beneficial for reviewers in this age group (e.g. Lu & Law, 2012; Wu & Schunn, 2021). However, relatively few studies have explored giving feedback to peers at this level of schooling. Another difference from most of the previously-cited studies which examined either the quality of the feedback provided or the reviewers' own learning products, the current study took a broader look at the reviewers' learning. In our study, learning was analyzed through several parameters: post-test scores, the quality of students' own concept maps, and the quality of the feedback given. Finally, our study was conducted in an online inquiry learning environment which, by its nature, provided a specific context for giving feedback to peers.

Overall, this led to the following research questions:

- Does the quality of reviewed products affect reviewers' learning when giving feedback to peers in a secondary-school inquiry learning context?
- Is there a differential effect of the level of quality of the reviewed product on reviewers' learning for students of different prior knowledge levels?

Based on the above literature, we expected that the level of reviewed products would affect reviewers' learning, with higher-quality products leading to more learning, and that students with higher prior knowledge could learn more from reviewing, because their knowledge allows them to understand and comment on products of different levels.

## Method

### Participants

The sample originally included 95 students from four classes at the same educational level (higher general secondary education, HAVO in Dutch) from one secondary school in the Netherlands, with average age of 15.21 years ( $SD=0.48$ ). For the 78 students (36 boys and 42 girls) who participated in all three sessions and gave feedback to peers, only their test results, concept maps and feedback given were used for the analysis. We excluded one participant during the preliminary data analysis, which gave a final sample of 77 students (see the Results section for details). Within each class, students were assigned to one of three conditions:

- giving feedback on a set of concept maps of low quality (condition 1).
- giving feedback on a set of concept maps of mixed quality (condition 2).
- giving feedback on a set of concept maps of high quality (condition 3).

The aim was to distribute prior knowledge between the conditions as similarly as possible. The scores for the pre-tests were used to determine prior knowledge (see the Domain knowledge test section for the details). In each class, students were first ranked according to their pre-test scores and then assigned to condition: the first, fourth, and so on students were assigned to condition 1; the second, fifth, and so on students were assigned to condition 2; and the third, sixth, and so on students were assigned to condition 3. In the final sample, there were 29, 23 and 25 students in conditions 1, 2 and 3, respectively.

## Study design

Students were asked to give feedback on a set of two concept maps, differing in condition, by fictitious students. The first set (condition 1) included two concept maps of low quality, with a very limited number of concepts used and a misconception. The second set (condition 2) consisted of two concept maps of mixed quality—one low-quality map (from the first set) and one high-quality map (from the third set). Finally, the third set included two concept maps of higher quality with more concepts used, but also still a misconception. This means that each set included at least one misconception.

Students in all conditions gave feedback by responding to criteria-related questions, that explicitly provided students with the important features of the reviewed product and implicitly pointed to the desired state of the product. For example, the question “What important concepts are missing?” explicitly asked students to name them and implicitly meant that a good concept map should include essential concepts related to the topic. In this way, while going through the questions, students were also encouraged to think about the important characteristics of a concept map. This approach to assisting and guiding students through the peer-feedback process was based on studies suggesting a model of this process. For example, a model developed by Sluijsmans (2002) includes providing students with assessment criteria as a starting point of the peer-feedback process.

The assessment criteria used in the study were neither domain- nor subject-dependent. Instead, they were specific to a particular type of a product (in our case, a concept map) and described its important characteristics. The criteria used were based on a study by van Dijk and Lazonder (2013) and are presented in the Table 1. For ethical reasons (to avoid any dependency), we did not share test results and created learning products with the teachers at the individual (student) level.

## Materials

The lesson was given in the students’ native language (Dutch) and covered a topic from the national physics curriculum, namely, the states of matter. In particular, it focused on the three basic states of matter (gas, liquid, solid) and the processes of transition between them.

Students worked with an online inquiry learning environment that was created using the Go-Lab ecosystem (see [www.golabz.eu](http://www.golabz.eu)). Inquiry learning imitates the scientific research cycle and leads students through a (guided) exploration process. The lesson was constructed as an inquiry learning space (ILS) that facilitated students’ working through an inquiry learning cycle, with experimenting in a virtual laboratory as the pivotal activity (de Jong et al. 2021).

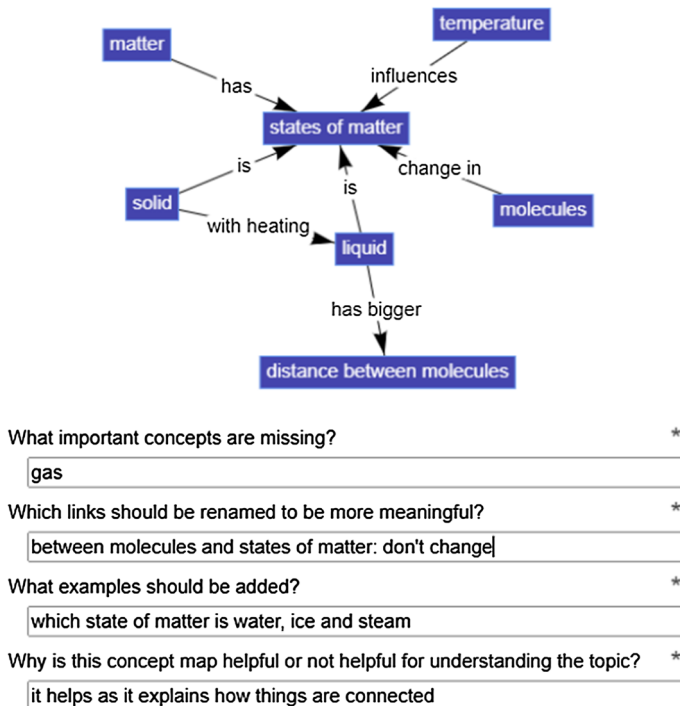
**Table 1** Assessment criteria questions for giving feedback

Question	Question wording
1	What important concepts are missing?
2	Which links should be renamed to be more meaningful?
3	What examples should be added?
4	Why is this concept map helpful or not helpful for understanding the topic?

The ILS followed the inquiry learning cycle described by Pedaste et al. (2015) and included several stages: orientation, conceptualization, experimentation, conclusion and discussion. During these stages, students created several types of learning product, such as hypotheses, concept maps, designs for experiments, and conclusions. The final stage—discussion—was used to reflect on the process and the learning products that had been created, which made this stage important for deeper learning. The activity of giving feedback to peers was positioned in the final stage and gave students an opportunity to evaluate (fictitious) peers' concept maps and revisit their own concept maps. A more detailed description of the ILS is given in the Appendix.

## Feedback tool

Using a special peer assessment tool, students gave feedback on two concept maps that were created by the research team. This tool presented students with a concept map by a (fictitious) peer student and a set of assessment criteria (Fig. 1) which were formulated in question form (see Table 1) so that students could give their feedback by answering these questions. Students were encouraged to give thoughtful feedback that could help these peers to improve their work. They were told that the concept maps had been created by some other students, who were not necessarily from the same class. The process of giving feedback was anonymous for students.



**Fig. 1** View of feedback tool with a higher-quality concept map (translated into English) and an aggregated example of feedback given by students

## Concept maps

In total, four concept maps with different levels of quality were created for the study. All concept maps were arranged in sets, with a different set for each condition. Set 1 (condition 1) consisted of two lower-quality concept maps, set 2 (condition 2) included one concept map from the lower-quality set and one concept map from the higher-quality set, and set 3 (condition 3) consisted of two higher-quality concept maps. Lower-quality concept maps had fewer concepts than higher-quality ones (four versus six or seven, respectively), but all concept maps lacked some important concepts. Each set of concept maps included at least one misconception; the covered misconceptions were that molecules change when going to another state of matter and that the distance between molecules influences the temperature of the matter. For example, one of the higher quality concept maps is presented in Fig. 1. This concept map lacked the concept of ‘gas’ as one of the phases and the misconception was that molecules change when going to another state of matter.

## Domain knowledge test

Parallel tests were used for pre- and post-testing, with questions focusing on understanding of phase-transformation processes. Each test consisted of six open-ended questions, with the maximum total score of 13 points: five questions worth two points and one question worth three points; the number of points represented the number of expected elements in the answer. Examples of questions and the grading scheme are presented in Table 2.

## Procedure

Students’ prior domain knowledge was assessed during the physics lesson just before the lesson with the ILS, using the pre-test (15 min) in an online form. For the lesson with the ILS, students in each class were assigned to conditions to make the distribution of the pre-test scores between the conditions as similar as possible.

Because the content of the lesson with the ILS was part of the standard school curriculum, the lesson was delivered during a regular physics lesson. The research gave a brief introduction about the procedure at the beginning of the lesson.

Students worked with a version of the learning environment that matched their assigned condition for one whole lesson (50 min). The researcher was present during the lesson and participants could ask questions about the procedure or tools, but not about the content. Students worked individually, using their computers to go through the lesson material at their own pace. They could go back and forth freely between the stages of the ILS. The first four stages of the ILS (orientation, conceptualization, investigation, and conclusion) focused on the inquiry learning activities. The final stage (discussion) covered the process of giving feedback on peers’ concept maps, together with the opportunity to improve students’ own concept maps (created earlier during the lesson, but presented again in this final stage) after that.

At 25 min and 40 min after the beginning of the lesson, students were reminded about the time and advised to proceed in their work with the material; however, this was not an obligation.

**Table 2** Example of the grading scheme for test answers

Question	Answers	Points
Do molecules change when changing to a different phase? Why?	No, they do not	1
	No, they do not, because the substance is the same	2
Describe the process of condensing. Think of energy and different characteristics of molecules	Smaller distance between molecules	1
	Smaller distance between molecules and molecules move at a lower speed	2
	Smaller distance between molecules, molecules move at a lower speed and the temperature of the substance is lower	3



At the beginning of the following physics lesson, the parallel post-test (15 min) was administered online.

## Analysis

To answer the research questions, the following statistical tests were used:

1. A repeated measures ANOVA to check that students had learned during the experiment.
2. An ANOVA with prior knowledge level and condition as independent variables and post-test scores as the dependent variable, with the aim of answering the Research question 1 (the effect of condition on reviewers' learning) and Research question 2 (the differential effect of prior knowledge in connection with the condition on reviewers' learning).
3. A MANOVA with condition as independent variable and characteristics of students' own concept maps and the feedback given as the dependent variables, with the aim of answering the Research question 1 (the effect of condition on reviewers' learning).

The researcher scored pre- and post-tests. A subset (41% of the pre-tests and 28% of the post-tests) were graded independently by a second rater (outside the research team), with Cohen's kappa being 0.83 and 0.80, respectively. We scored tests using a scheme covering what the answer should include to obtain the points. The scheme was developed by the researchers and approved by teachers at the participating school.

Students' pre-test scores were used as an indicator of their prior knowledge. To answer the question about the differential effect of condition for different prior knowledge groups, students from the whole sample were divided into three groups: low prior knowledge (pre-test score more than 1SD below the mean), average prior knowledge (pre-test score within 1SD above or below the mean) and high prior knowledge (pre-test score more than 1SD above the mean).

Students' final versions of their concept maps (after interaction with the learning environment, providing feedback, and optional revision of the concept map) were also assessed by the researchers. The following characteristics of the concept maps were evaluated: proposition accuracy score (the number of correct links), salience score (the proportion of correct links out of all links), and complexity score (the level of complexity).

A general representation of the main concepts and links between them was created together with an expert, and this was used as a reference source for coding students' concept maps. All three scores were used separately in the analysis. Proposition accuracy had no maximum score, because students could present as many as they wanted, provided that the concepts were relevant. Because students did not receive any specific training on a particular way to do concept mapping, any approach was considered viable; therefore, no comparison was made to a criterion map.

The salience score had a maximum of 1. The complexity scoring used a nominal scale, with the goal of discriminating between a linear representation ('sun' or 'snake' shapes) and a more-complex structure with more than one hierarchy level, with 1 point assigned for a linear structure and 2 points assigned for a more complex structure. A subset (19%) of the concept maps were coded by a second rater, with Cohen's kappa of 0.70.

Finally, we assessed the quality of the feedback given. Correct answers for the criteria questions were worth 1 point each, while each additional meaningful suggestion for the same criterion gave them 0.5 points. A subset (25%) of the feedback was coded by a

second rater, with Cohen's kappa reaching 0.71. An aggregated example of the type of feedback given by students is presented in Fig. 1.

For a deeper understanding of the reviewers' learning process, we planned to use an exploratory regression analysis. It zoomed in on the role that the coded characteristics of students' own concept maps and the quality of the feedback given play in explaining the post-test scores.

## Results

During the preliminary analysis of the data, one participant was found with a difference score between post-test and pre-test that was more than 2.5 *SD* from the mean difference score for the whole sample. In particular, the student scored 8 (out of 13) on the pre-test and 1 (out of 13) on the post-test. It seemed very improbable that a student could lose so much knowledge within a week. Therefore, we labeled this participant as an outlier and excluded this participant's data from the analysis.

The difference in pre-test scores between the conditions was not significant:  $M_1 = 3.52$  ( $SD = 1.60$ ),  $M_2 = 3.91$  ( $SD = 1.73$ ),  $M_3 = 3.80$  ( $SD = 2.06$ ),  $F(2,74) = 0.34$ ,  $p = 0.71$ . To check whether students did learn, a repeated measures ANOVA was used to compare pre- and post-test scores, with test scores as a within-subjects factor and condition as a between-subjects factor. There was a significant effect (with a large effect size) of the experimental lesson, indicating that students did learn both as a whole group [Wilks' lambda = 0.73,  $F(1,76) = 27.83$ ,  $p < 0.001$ ,  $\eta^2 = 0.27$ ] and in each condition. For condition 1, Wilks' lambda = 0.64,  $F(1,28) = 15.50$ ,  $p < 0.001$ ,  $\eta^2 = 0.36$ ; for condition 2, Wilks' lambda = 0.73,  $F(1,22) = 8.10$ ,  $p < 0.05$ ,  $\eta^2 = 0.27$ ; and for condition 3, Wilks' lambda = 0.73,  $F(1,24) = 5.65$ ,  $p < 0.05$ ,  $\eta^2 = 0.19$ .

**Table 3** Mean test scores by knowledge level and condition (maximum score is 13)

Condition	Low prior knowledge		Average prior knowledge		High prior knowledge		Total	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Condition 1	<i>(n = 8)</i>		<i>(n = 17)</i>		<i>(n = 4)</i>		<i>(n = 29)</i>	
Pre-test	1.50	0.76	3.88	0.70	6.00	0.00	3.52	1.56
Post-test	5.75	2.96	5.18	2.66	8.25	0.96	5.76	2.74
Condition 2	<i>(n = 7)</i>		<i>(n = 11)</i>		<i>(n = 5)</i>		<i>(n = 23)</i>	
Pre-test	2.00	0.00	4.00	0.78	6.40	0.89	3.91	1.73
Post-test	3.00	1.63	5.55	1.63	6.60	2.41	5.00	2.24
Condition 3	<i>(n = 8)</i>		<i>(n = 11)</i>		<i>(n = 6)</i>		<i>(n = 25)</i>	
Pre-test	1.75	0.46	3.64	0.81	6.83	0.98	3.80	2.06
Post-test	3.13	2.90	4.91	2.02	7.00	1.27	4.84	2.58
Total	<i>(n = 23)</i>		<i>(n = 39)</i>		<i>(n = 15)</i>		<i>(n = 77)</i>	
Pre-test	1.74	0.54	3.85	0.75	6.47	0.83	3.73	1.72
Post-test	4.00	2.81	5.21	2.20	7.20	1.70	5.23	2.54

## Test scores (Research questions 1 and 2)

Regarding the effect of condition and prior knowledge on students' test results, descriptive statistics for participants' test results are presented in Table 3. ANOVA was used to answer the research questions about the effect of the reviewed products' quality on reviewers' post-test scores, with post-test score as a dependent variable and prior knowledge level and condition as independent variables (between-subjects factors). Only the main effect for prior knowledge was significant, with a large effect size [ $F(2, 68)=9.51, p<0.01, \eta_p^2=0.22$ ] showing that higher prior knowledge led to higher post-test scores across conditions. Similar results were found using a continuous variable for prior knowledge (pre-test scores), which reflects the original procedure better than a categorical variable: a statistically-significant effect of prior knowledge with a large effect size [ $F(8, 56)=2.22, p<0.05, \eta_p^2=0.24$ ].

Inspection of the results of the follow-up analysis showed that the results for condition 1 (lower quality concept maps) differed somewhat from those for the two other conditions, especially for the low prior knowledge group. The trends can be seen in Fig. 2. A pairwise comparison of the estimated marginal means of post-test scores between the conditions showed a statistically-significant difference between condition 1 and condition 3 ( $p=0.048$ ). Students from condition 1 (lower-quality concept maps) were found to have higher estimated post-test scores than students from condition 3 (higher-quality concept maps) [ $M_1=6.39, SE=0.50; M_3=5.01, SE=0.47$ ].

## Concept map and feedback quality (research question 1)

MANOVA was conducted to check if the characteristics of students' own concept maps and the feedback given differed by condition. Descriptive statistics are presented in Table 4. Results for the concept map characteristics showed no significant differences between conditions: Wilks' lambda=0.79,  $F(10, 124)=1.55, p=0.13$  [ $F_{\text{CHANGED}}(2, 66)=0.46, p=0.63; F_{\text{ACCURACY}}(2, 66)=0.79, p=0.46; F_{\text{SALIENCE}}(2, 66)=0.32, p=0.73; F_{\text{COMPLEXITY}}(2, 66)=1.01, p=0.37$ ]. As for feedback quality, the highest-scoring feedback, on average, was given by the students who evaluated the low-quality set of concept maps (condition 1). Students who reviewed the mixed set of concept maps (condition 2)

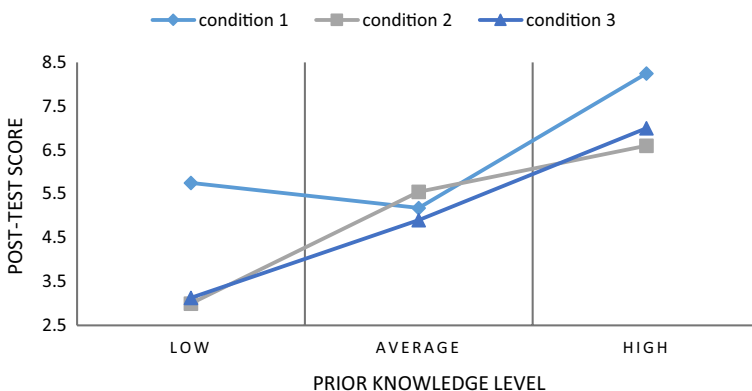


Fig. 2 Marginal means of post-test scores (estimated) for each prior knowledge group and condition

**Table 4** Mean scores for concept maps and feedback quality by condition

CM characteristics and feedback quality	Maximum score	Condition 1 (low-quality CMs)		Condition 2 (mixed CMs)		Condition 3 (high-quality CMs)	
		<i>(n</i> = 26)		<i>(n</i> = 23)		<i>(n</i> = 20) <sup>1</sup>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Changed CM after giving feedback	1.00	.23	.43	.17	.39	.30	.47
CM proposition accuracy	n/a	1.85	2.33	2.13	2.38	2.80	3.11
CM salience score	1.00	.40	.45	.51	.49	.46	.47
CM complexity	2.00	0.88	0.59	0.70	0.47	0.90	0.55
Feedback quality	n/a	2.43	1.07	1.82	0.90	2.03	0.98

CM = concept map

<sup>1</sup>Due to technical reasons, the feedback provided by eight students in this condition was not saved, so their data were not included in this analysis

gave the lowest-scoring feedback, on average. Moreover, pairwise comparison showed that the difference in the quality of feedback between these two conditions (condition 1 and 2) was statistically significant, with students from condition 1 providing higher-quality feedback than students from condition 2 [ $M_1 = 2.43$ ,  $SE = 0.19$ ;  $M_2 = 1.82$ ,  $SE = 0.21$ ,  $p = 0.033$ ].

For an exploratory regression on the role of students' produced products in their learning, the following variables were used: the fact of changing one's own concept map after giving feedback, the number of correct propositions (proposition accuracy score); the proportion of correct propositions (salience score), the complexity of the concept map (complexity score); and the quality of the feedback given. A forward regression was used to check which variables predicted the post-test score. Changing the student's own concept map was found to be a significant predictor, indicating that students who had changed their concept maps scored higher on the post-test ( $B = 2.01$ ,  $\beta = 0.34$ ,  $p = 0.005$ ). Because the focus of attention was on characteristics of the learning products, prior knowledge was not included in this analysis. However, adding it to the forward regression gave similar results: both prior knowledge ( $B = 1.35$ ,  $\beta = 0.38$ ,  $p = 0.001$ ) and changing concept maps ( $B = 1.64$ ,  $\beta = 0.27$ ,  $p = 0.013$ ) were significant predictors of post-test scores.

## Conclusions and discussion

The focus of this study was the roles played in the learning of a peer reviewer by the quality of products that were reviewed and, in relation to that, the prior knowledge of students. There can be different ways of measuring reviewers' learning: post-test scores, the quality of students' own products (of the same type as the reviewed product), or the quality of feedback provided. While the cited studies used either one of these ways or a combination of several, the current study operationalized learning as involving all three components, which made comparison with previous research results possible.

Based on the literature (e.g. Alqassab et al., 2018a; Tsivitanidou et al., 2018), we expected that reviewing higher-quality products would be more beneficial for reviewers' learning, because they would see a good concept map and understand the topic better. We had this expectation even though some of the studies that we consulted involved

implementing reciprocal peer assessment, which meant that the feedback received by reviewers could also influence the results. Another expectation was that students with higher prior knowledge would benefit more from the reviewing process, because their level of knowledge allowed them to provide more criticism, thereby being more cognitively involved (e.g. Cho & Cho, 2011; Patchan & Schunn, 2015). The findings, however, showed opposite trends.

According to the results, giving feedback on the set of lower-quality concept maps led to higher post-test scores for the reviewers than giving feedback on the set of higher-quality concept maps. Even though this finding is opposite from what was expected, it can still be explained by theory. One explanation might be connected with participants' prior knowledge level. Most researchers have argued that students should be competent enough in the domain to give meaningful and useful feedback which, in turn, can lead to their own learning (e.g. Alqassab et al., 2018b). In our case, students might not have been very knowledgeable about the topic even after the lesson, because the mean overall post-test score was just about 5 out of 13 points. This could mean that lower-quality concept maps matched their knowledge level better, and so they could give meaningful feedback on this level of products. According to the trends, reviewing lower-quality concept maps was especially beneficial for the group of students with lower prior knowledge, which supports our argument. For practice, this might mean that students should be grouped in order to give feedback on peers' products of approximately the same level as theirs or lower. In the case of an online learning environment, a teacher can configure the collaboration within the environment in a way that allows students of similar levels of knowledge to give feedback to each other.

Analysis of the feedback given supports the importance of having enough knowledge (and being confident enough) to give feedback. The highest scored feedback was given by the group reviewing lower-quality concept maps and the lowest scored feedback was provided by the group reviewing the mixed set. For the mixed set, the feedback on the higher-quality concept map was worse than on the lower-quality concept map. It was more difficult for students to identify mistakes and provide suggestions for a higher-quality product. This might be explained by students' natural comparison of the two concept maps to each other, even though there was no such task. Because students saw the lower-quality concept map first, the second concept map could have seemed so much better that they did not spot the mistakes. Finding only a few mistakes in a generally high-quality product could be more difficult than finding obvious mistakes in a generally low-quality product.

To summarize these two types of results, it appeared that the most learning happened for students who reviewed a set of lower-quality concept maps (condition 1)—both their post-test scores and the quality of their feedback were higher. Because this was true for students with different prior knowledge levels, having enough knowledge to evaluate the presented product leads to being able to provide meaningful feedback and to better learning. In our case, even for higher prior knowledge students, giving feedback on higher-quality concept maps with just a misconception appeared to be too challenging. Evaluating a concept map was a challenging task because concept mapping itself was a new activity for students. Evaluating a concept map with obvious flaws was therefore a more-feasible task. This is supported by the findings by van Zundert et al. (2012) who demonstrated that task complexity has a more-negative effect on peer assessment skills than on domain-specific skills, because peer assessment requires more higher-order thinking and skills. In other words, having the more-challenging task of finding mistakes in a higher-quality concept map led to lower-quality feedback (fewer mistakes spotted, fewer suggestions made) which, in turn, could lead to less learning. For practice, this might mean that, to maximize learning,

students should become familiar with the product to be reviewed before they review it, especially with complex products.

Better learning could also be connected to the fact that students could provide more-critical feedback on the set of lower-quality concept maps. According to Yalch et al. (2019), providing specific critical feedback leads to learning. Critically processing a peer's product means more cognitive involvement with the material which, in turn, is reflected in post-test scores. This argument is supported by the finding that changing students' own concept maps was a significant predictor of their post-test results. Students who re-worked their concept maps after giving feedback applied their thinking about the peer's product to their own product and, as a result, outperformed other students on the post-test. For practice, this can mean that including a required activity of reworking students' own products after they have given feedback could increase their learning.

Another explanation of the difference between the expected and obtained results could be the differences in the reviewed products. All of the studies used for the literature review invited students to assess bigger and more-complex learning products—from essays and reports to geometry proofs (e.g. Alqassab et al., 2018a; Cho & Cho, 2011; Patchan & Schunn, 2015; Tsivitanidou et al., 2018), while the current study focused on a very compact product—a concept map. On the one hand, a concept map is more structured and, thus, mistakes regarding to missing parts can be more obvious, which makes it easier for weaker students to give feedback because they can identify those mistakes. On the other hand, creating a concept map as an exercise requires much more abstract thinking, which is generally more difficult for students. Moreover, approaches to concept mapping can be quite different for different students, and understanding a different rationale used in another's concept map also can be rather difficult. Therefore, spotting mistakes in higher-quality concept maps could be too challenging even for students with higher prior knowledge. These two factors together made concept maps of lower quality more preferable for students with different levels of prior knowledge.

Our study was conducted in a specific context (inquiry learning) and with quite specific small products (concept maps), both of which were different from most of the cited studies. This could limit the generalizability of the findings and interpretation. Another limitation could be the number of participants, which was just sufficient for a study with three conditions. Having a larger group of participants might make the differences between the conditions more obvious. Therefore, one direction for further research could be checking the results of this study with a larger sample.

## Concluding notes

Several more general conclusions can be drawn based on the results of this study. First, students' prior knowledge should be taken into account when grouping them for giving feedback. In particular, when dealing with smaller learning products, it might be more beneficial for students' learning to have them review pieces of work that are at or (a bit) below their current level of knowledge. This would help them to associate themselves with those products and to feel that they have enough knowledge to provide meaningful feedback. This could lead to more learning than giving students a higher-quality product that they would not be able to analyze and learn from. Seemingly, weaker learning products stimulate students to provide more feedback that is critical and, as a result, students learn more. Second, students who had reworked their own concept maps scored higher on the post-test, suggesting that taking another step after giving feedback, namely, revisiting one's own

product, could contribute to learning. Therefore, students should be encouraged to revise their own products after providing feedback. Finally, to establish the generalizability of our results, future research could check whether these findings hold up for other small learning products created during the inquiry learning process, such as hypotheses or conclusions.

## Appendix

### Detailed description of the online learning environment used during the experiment

Stages of the ILS:

- *Orientation* — In this stage, the topic was introduced, as well as the idea of doing scientific research through experimenting in an online laboratory. The research question about the mechanism of phase transitions was set, and students' prior knowledge of the topic was refreshed with a brief quiz covering the names of some phase changes and examples of them from daily life.
- *Conceptualization* — In this stage, the topic was narrowed to questions about the process of phase changing. Students could think about the most-important concepts related to the topic by creating a concept map using a specially-designed tool, the Concept Mapper (see Fig. 3). The tool included some pre-defined concepts for the topic and names of links, organized in a drop-down menu. It also gave students the possibility of typing their own concepts and names of links. This combination provided some guidance, but still allowed freedom of learning. At the end of this stage, students were

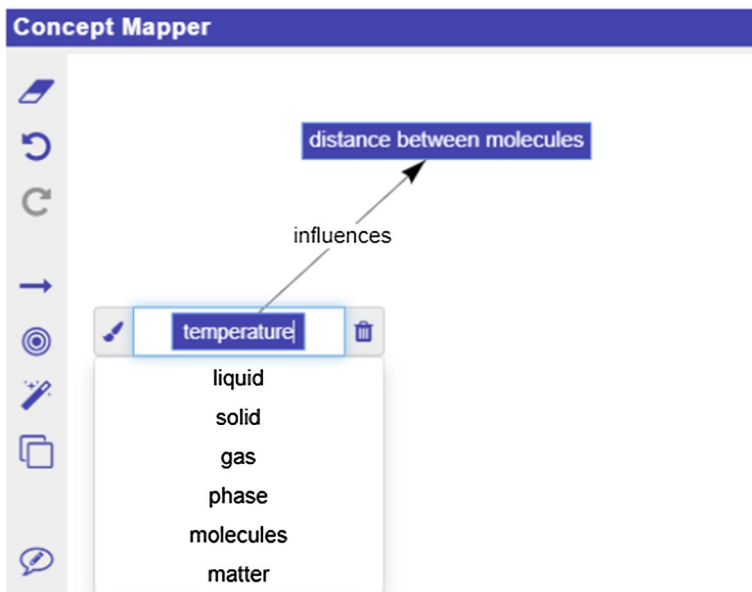
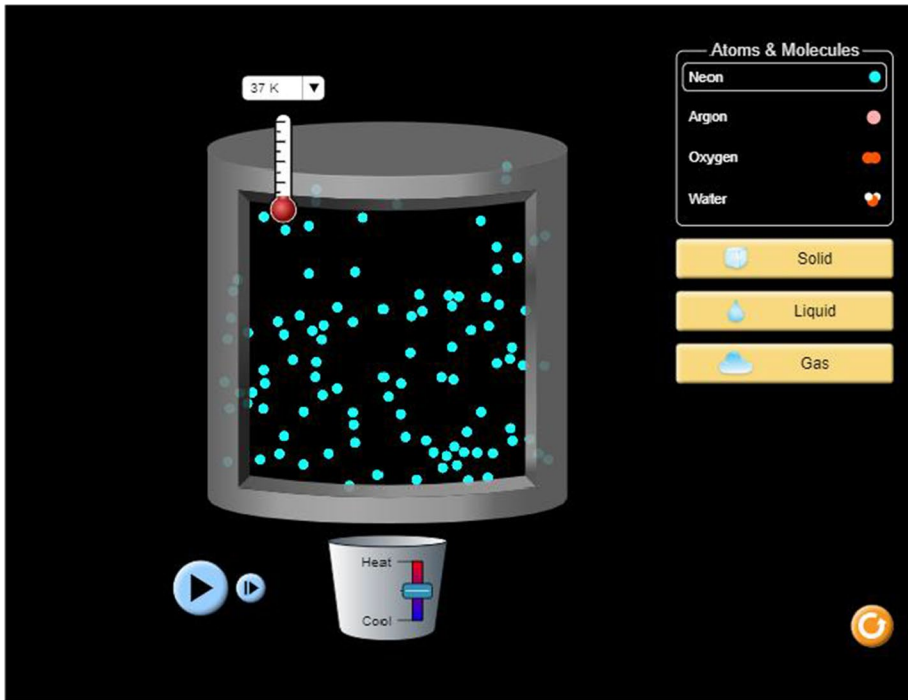


Fig. 3 View of the concept mapper tool (translated into English)



**Fig. 4** 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>

encouraged to create hypotheses about changes in matter that would be noticeable during a phase change. This hypothesis-creation process was supported by another tool (Hypothesis Scratchpad) with some pre-defined parts. As with the Concept Mapper, students could use pre-defined parts as well as typing in their own terms to write hypotheses.

- *Investigation* — In this stage, the virtual laboratory (see Fig. 4) was used for conducting an experiment and checking hypotheses. A brief introduction to the laboratory's settings was provided in the ILS, but students had the freedom to design the experiment themselves. In the laboratory, they could change the temperature of the matter to see what happened with its molecules. They could also draw correspondences between different stages of the temperature-changing process and the states of matter – solid, liquid, gas. Observations could be noted down in a special input box.
- *Conclusion* — In this stage, the conclusion was to be developed, based on their hypotheses and observations in the online laboratory. Students were shown the hypotheses that they had created before the experiment and were asked to accept or reject them based on their observations, as well as to explain their choice.
- *Discussion* — In this stage, students were supposed to reflect on their created learning products and discuss their experience. In our study, students were asked to give feedback on two pre-defined concept maps. This process was supported by a specially-developed feedback tool (see Fig. 1). After this activity, students were invited to rework



their own concept maps (created in the conceptualization stage) and make any necessary changes. Students did not have to go back to the Conceptualization stage to do this; instead, their original concept maps could be downloaded to make the process more natural.

**Acknowledgements** 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. 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. 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.

**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

- Alqassab, M. (2017). *Peer feedback provision and mathematical proofs: Role of domain knowledge, beliefs, perceptions, epistemic emotions, and peer feedback content*. Unpublished doctoral dissertation, Ludwig-Maximilians University, München, Germany.
- Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018a). 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>
- Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018b). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education*, 33, 11–30. <https://doi.org/10.1007/s10212-017-0342-0>
- Brill, J. M. (2016). Investigating peer review as a systemic pedagogy for developing the design knowledge, skills, and dispositions of novice instructional design students. *Educational Technology Research and Development*, 64, 681–705. <https://doi.org/10.1007/s11423-015-9421-6>
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39, 629–643. <https://doi.org/10.1007/s11251-010-9146-1>
- de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., Dziabenko, O., Koslowsky, J., Korventausta, M., Law, E., Pedaste, M., Tasiopoulou, E., Vidal, G., & Zacharia, Z. C. (2021). Understanding teacher design practices for digital inquiry-based science learning: The case of Go-Lab. *Educational Technology Research & Development*, 69, 417–444. <https://doi.org/10.1007/s11423-020-09904-z>
- Dmoshinskaia, N., Gijlers, H., & de Jong, T. (2021). Learning from reviewing peers’ concept maps in an inquiry context: Commenting or grading, which is better? *Studies in Educational Evaluation*, 68, 100959. <https://doi.org/10.1016/j.stueduc.2020.100959>
- 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>
- 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>
- Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40, 257–275. <https://doi.org/10.1007/s11251-011-9177-2>

- Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2014). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, 39, 657–677. <https://doi.org/10.1080/02602938.2013.860421>
- Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, 41, 381–405. <https://doi.org/10.1007/s11251-012-9236-3>
- 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., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & 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>
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment. The training of peer assessment skills*. Unpublished doctoral dissertation, Open University of the Netherlands.
- Tenório, T., Bittencourt, I. I., Isotani, S., & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior*, 64, 94–107. <https://doi.org/10.1016/j.chb.2016.06.020>
- Tsvitanidou, 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/s10212-017-0341-1>
- Tsvitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21, 506–519. <https://doi.org/10.1016/j.learninstruc.2010.08.002>
- 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*, 24, 194–204. <https://doi.org/10.1080/10494820.2013.834828>
- van Zundert, M. J., Sluijsmans, D. M. A., Könings, K. D., & van Merriënboer, J. J. G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology*, 32, 127–145. <https://doi.org/10.1080/01443410.2011.626122>
- Wu, Y., & Schunn, C. D. (2021). 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. <https://doi.org/10.3102/0002831220945266>
- Yalch, M. M., Vitale, E. M., & Ford, J. K. (2019). Benefits of peer review on students' writing. *Psychology Learning & Teaching*, 18, 317–325. <https://doi.org/10.1177/1475725719835070>

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