Giving Feedback on Peers' Concept Maps in an Inquiry Learning Context: The Effect of Providing Assessment Criteria

Natasha Dmoshinskaia¹ · Hannie Gijlers¹ · Ton de Jong¹

Accepted: 8 November 2020 / Published online: 18 November 2020
© The Author(s) 2020

Abstract

Giving and receiving peer feedback is seen as an important vehicle for deep learning. Defining assessment criteria is a first step in giving feedback to peers and can play an important role in feedback providers' learning. However, there is no consensus about whether it is better to ask students to think about assessment criteria themselves or to provide them with ready-made assessment criteria. The current experimental study aims at answering this question in a secondary school STEM educational context, during a physics lesson in an online inquiry learning environment. As a part of their lesson, participants ($n = 93$) had to give feedback on two concept maps, and were randomly assigned to one of two conditions—being provided or not being provided with assessment criteria. Students' post-test scores, the quality of feedback given, and the quality of students' own concept maps were analyzed to determine if there was an effect of condition on feedback providers' learning. Results did not reveal an advantage of one condition over the other in terms of learning gains. Possible implications for practice and directions for further research are discussed.

Keywords Peer assessment · Giving feedback · Inquiry learning · Assessment criteria · Concept maps

Introduction

Peer assessment is being used more and more in education. Its growing popularity is due in part to the trend of making educational processes in general, and assessment processes in particular, more active and student-centered (de Jong 2019). This increasing popularity is backed up by empirical evidence; a recent meta-analysis (Li et al. 2020) showed that students involved in peer assessment have higher learning performance than those not involved.

Peer assessment actually includes two processes: giving feedback to peers and, in turn, receiving feedback from them. Current research on peer assessment is still focused more on the comparison of peer and teacher assessments and on the effects of receiving the feedback, and less on the students' learning that results from giving feedback. However, several studies have demonstrated that students view

giving feedback as a more valuable learning experience than receiving peers' feedback and that feedback providers have better learning results than feedback receivers (e.g., Cho and MacArthur 2011; Ion et al. 2019).

The present study targets this less investigated area of giving feedback to peers. According to the model suggested by Sluijsmans (2002), giving feedback includes the following steps: define assessment criteria, judge the performance of a peer, and provide feedback for future learning. Following the approach of decomposing the process in order to develop better understanding, we focused on the first step, defining assessment criteria, and the effect that this has on the performance of feedback providers, whom we refer to as reviewers. More specifically, we investigated whether being provided with criteria for peer assessment affects reviewers' learning.

Some work has indicated that having students define their own criteria for giving feedback is more beneficial than providing them with ready-made criteria. According to a review on peer assessment by Falchikov (2004), familiarization with assessment criteria and the related feeling of ownership lead to higher feedback validity (correspondence between peer and expert feedback). Such familiarization makes students feel more responsible and self-regulated in their learning.

✉ Natasha Dmoshinskaia
n.dmoshinskaia@utwente.nl

¹ Department of Instructional Technology, Faculty of Behavioral, Management, and Social Sciences (BMS), University of Twente, P.O. Box 217, 7500, AE, Enschede, The Netherlands

The results of the study by Canty et al. (2017) showed that thinking of their own assessment criteria gives students more responsibility for their learning, which, in turn, results in more involvement in the assessment process. In their study, students had to reflect on the important characteristics that the assessed product should possess, which contributed to their development of domain knowledge and assessment skills. These authors reported that students reacted positively to the absence of explicit assessment criteria, and were also more engaged with the task of giving feedback when no criteria were provided. Moreover, a study by Tsivitanidou et al. (2011) showed that students not only reacted positively towards giving peer feedback without provided assessment criteria, but could also come up with their own criteria; in other words, they could perform basic assessment tasks themselves.

An alternative position on the value of providing assessment criteria is based on the understanding that, in general, the process of giving feedback is a challenging task for students; the need to develop assessment criteria from scratch can make that process too difficult and, thus, no longer useful for learning. Several studies (e.g., Gielen and De Wever 2015; Sluijsmans et al. 2002) have argued that students should be guided through the process of giving feedback, and that a more structured template for assessing peers improves the content of the feedback provided. This is supported by the findings by Gan and Hattie (2014), who studied the peer feedback process with secondary school students. In their experiment, students had to give feedback on their peers' chemistry reports and were divided into two conditions, with and without assessment prompts. Prompts were formulated as questions asking the student to evaluate strong and weak aspects of the report and to suggest ways to improve it. The questions stimulated students' thinking about the important characteristics of the report and how to include them. The group with the prompts (questions) produced higher quality feedback, spotting more errors and providing more suggestions than the group without prompts. Moreover, their own written products showed better quality compared to those of students from the group without prompts. It is possible that having or not having prompts triggered different learning processes, but the published results did not give a clear answer to this. The finding by Gan and Hattie (2014) that students from a prompt condition produced higher quality learning products aligns with the findings by Panadero et al. (2013). In their study, students had to create concept maps about teaching methods and gave feedback on the concept maps of their peers. Participants who were provided with assessment rubrics for giving feedback created higher quality concept maps than did participants in the condition without rubrics.

Although these studies underline the importance of providing students with assessment criteria, a review by

Deiglmayr (2018) demonstrated that just providing criteria might not be enough to improve reviewers' own learning performance. A conclusion from this review was that it is crucial for peer reviewers to have enough knowledge about the subject and enough understanding of the assessment criteria to benefit from giving feedback to their peers. Along this line, Peters et al. (2018) found that students from a condition with formative assessment scripts generated more and better quality feedback for an assessed draft by a fictitious peer, but that those students' own drafts were not of higher quality. This means that lessons learned from assessing someone else's work do not necessarily transfer to improvement of one's own products.

The above overview of studies shows that there is not a single clear picture that arises from the literature. The differences in outcomes from these studies could be caused by additional factors that influence the effect of providing students with assessment criteria or not. One such factor is the complexity of the topic. Jones and Alcock (2014) suggested that the absence of criteria helps when dealing with complex subjects, as the assessment criteria for such subjects might also be quite complex, creating extra difficulty. The authors suggested that it is easier to compare two pieces of work to each other than to gauge the quality of a piece of work against specified criteria. When students are not provided with assessment criteria, they compare the peer's work with their own vision of what this product should look like. In other words, they compare the peer's work with their own, which might be easier than comparing it with provided criteria. In the study by Jones and Alcock (2014), university students had to assess complex mathematical reports without any provided assessment criteria, and demonstrated high validity of such assessments. The authors argued that this validity could originate from the absence of assessment criteria, which would mean that not being provided with assessment criteria for more complex subjects can be better for reviewers. Accepting this view allows the contradiction between the findings by Jones and Alcock (2014) and Panadero et al. (2013) to be attributed to the difference between the domains involved: more complex (mathematics) and less complex (pedagogics). However, leaving students without provided assessment criteria for complex topics is contrary to the position taken by Könings et al. (2019). They stated that the complexity of the task can lead to poorer feedback in general, as dealing with two difficulties (completing the task and giving feedback) at the same time would create cognitive overload. This is especially important in situations where no assessment criteria are given to reviewers. Könings et al. (2019) based their approach on Bloom's taxonomy, pointing out that evaluating, particularly without guidance, can only be accomplished if the lower skills, such as applying and analyzing, have been mastered. This would mean that the more complex the domain is, the more support is

needed for reviewers, for example, in the form of providing assessment criteria.

A related, but slightly different factor that may influence the effect of providing or not providing assessment criteria is the knowledge that students bring to the task. The same topic can involve a different level of complexity for students with different levels of prior knowledge. Therefore, it is also possible that the benefits of having assessment criteria or not differ depending on students' levels of relevant prior knowledge. In their study, Orsmond et al. (2000) described what causes this difference. The authors believed that generating one's own criteria is less challenging than understanding and interpreting the teacher's criteria, as students do not go beyond their own understanding of the domain and the required skills when creating their own criteria. In other words, assessment criteria may be too abstract and difficult to understand for some students, so using self-created criteria would seem an easier option for them. Therefore, it is possible that more experienced students can better appreciate the challenge of using provided assessment criteria, while less-experienced students can benefit from the more comfortable situation of not having to deal with provided assessment criteria.

The process of giving feedback to peers can also be influenced by the context and the method by which it takes place. In the current study, giving feedback on peers' work was part of online inquiry learning. Inquiry learning allows students to explore a topic in a way that simulates scientific inquiry, which is a method well suited for teaching science. This also makes inquiry learning an even more natural context for giving feedback than a traditional lesson, as critiquing and peer feedback are part of a real research cycle, moreover, these activities develop scientific reasoning and promotes conceptual understanding (Dunbar 2000; Friesen and Scott 2013). Inquiry learning has also been found to activate students' involvement with the material and lead better learning results compared to traditional instructional methods (de Jong 2006; Minner et al. 2010). As for the online aspect, computer-facilitated methods of giving feedback were shown to be more beneficial for students' learning than paper-based methods, according to a recent meta-analysis by Li et al. 2020. This can be explained by the fact that online tools for giving feedback allow this process to be anonymous and done at a pace and time convenient for students, and increase their ownership of learning (Rosa et al. 2016).

To sum up, there is no clear opinion about the usefulness of providing students with assessment criteria for giving feedback to peers. On the one hand, granting students freedom concerning assessment criteria could stimulate their creativity, lead to deeper thinking about the product and give them more responsibility and ownership of their learning (e.g., Canty et al. 2017; Jones and Alcock 2014). On the other hand, the process of giving feedback could be challenging for students, and provided

assessment criteria can facilitate this process by orienting students to the desired state of the assessed product (e.g., Gielen and De Wever 2015; Sluijsmans et al. 2002). Another unclear aspect is whether the complexity of a topic for particular students or, in other words, their prior knowledge, plays a role in the possible effect of provided criteria. Finally, as the above-mentioned studies were not conducted in an inquiry learning context and not all of them used online tools, it is not clear yet if the results would be similar in this context.

This brings us to the following two *research questions*: What role does the presence or absence of provided assessment criteria play in reviewers' learning? Does the effect vary between groups with different levels of prior knowledge?

Method

Participants

The sample consisted of 137 8th grade students from two secondary schools in Russia, from three classes in each school. Participants had an average age of 14.27 years ($SD = 0.37$). Only students who were present for all three sessions (pre-test, the experimental lesson and post-test) and actually gave feedback were included in the analysis, which reduced the group to 93 students (54 girls and 39 boys). Out of 44 excluded students, 28 missed at least one session and 16 did not give feedback.

In each class, students were assigned to one of two conditions (with or without provided assessment criteria) in such a way that the distribution of prior knowledge in each condition was as similar as possible. Prior knowledge was determined based on pre-test scores. In the final sample of 93 participants, 44 were in the group without provided criteria and 49 in the group with provided criteria.

Study Design

Students were asked to give feedback on two concept maps. Participants in the free condition were not provided with any assessment criteria but were instead supposed to come up with assessment criteria themselves. Participants in the criteria condition were provided with assessment criteria in the form of questions. These questions directed students to consider the desired features of a concept map, and were not domain-specific. This is similar to the approach used by Gan and Hattie (2014), who used the word 'prompts' in this context. The assessment criteria for reviewing concept maps were based on the study by van Dijk and Lazonder (2013), and included the most important qualities of a concept map, such as the presence of main concepts, meaningful links, a hierarchical structure, and so forth. The questions used in the study as assessment criteria are presented in Table 1. Even though the assessment criteria were not domain-specific,

Table 1 Assessment criteria for the criteria condition (translated from Russian)

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 reviewed concept maps were, which led to giving domain-specific feedback.

The two concept maps that participants had to comment on were created by the research team. To make the setting realistic, participants were informed that these concept maps were created by other students, probably from another school. The use of concept maps by fictitious students meant that all students reviewed the same set of products, thereby enabling experimental control of that aspect.

The timeline of the experiment is presented in Fig. 1.

Materials

All instruction was given in the students' native language (Russian) and followed the national physics curriculum. The experimental lesson matched the school program and presented the planned curricular topic at the time of the experiment. The subject of the lesson was the physics topic of properties of matter, and more specifically concerned the three basic states of matter (solid, liquid, and gas) and transitions between them.

The lesson was constructed in an online inquiry environment, created with the help of the Go-Lab ecosystem (see www.golabz.eu). Such lessons present an inquiry learning space (ILS) that aims at facilitating students' inquiry process. During the lesson, students are led through an Inquiry Learning Cycle that consists of several phases, each having its own purpose (Pedaste et al. 2015).

The ILS used for the experiment followed a basic scenario for inquiry learning and included the following phases:

- Orientation—students were introduced to the topic and the idea of scientific experimentation in an online lab. The research question for the lesson was also set: how does matter change its state? Students' prior knowledge on the topic was activated with a short quiz (different from the pre-test).

- Conceptualization—students could express their ideas about the transitions between states of matter by creating a concept map. This was done with the help of a special tool (see Fig. 2) in which students could use some predefined concepts and names of links, as well as create their own concepts and links. Predefined concepts and names of links served as a scaffold for the concept-mapping process. After creating a concept map, students were asked to write down one or more hypotheses (with the help of a specific scaffolding tool) that they wanted to check in an online lab.
- Investigation—students conducted an experiments in an online lab (see Fig. 3) and wrote down their observations. Brief instruction on how to use the lab was given; however, students decided themselves how to conduct the experiment. The lab visualized the process of transition between different states of matters, as well as the influence of temperature on this process.
- Conclusion—students could compare their original hypotheses and observations made during working with the lab and use this information to answer the research question. Again, a specific online tool was available for this.
- Discussion—first, students gave feedback on two concept maps; then, they were asked to re-work their own concept maps that had been created in the Conceptualization phase.

The two concept maps that students reviewed differed in quality: one had only a few concepts that were not all linked with each other, while the other presented more concepts with more complex connections. Both concept maps were missing some concepts and included a misconception/mistake; in other words, both concept maps could be improved. For the first concept map, the mistake was that distance between the molecules influences the temperature and for the second map, the misconception was that molecules change during the transition between different states (see Figs. 4 and 5, respectively).

The decision to control the level of reviewed products was influenced by the possibility that the quality of the feedback and the reviewer's learning may depend on the quality of the learning products being reviewed (see, e.g., Patchan and Schunn 2015). The concept maps needing feedback were introduced to students in the same order: first the one of lower quality, then the one of higher quality. This order was chosen to prevent students from copying ideas from the better concept map in their suggestions for the lower quality concept map.

Fig. 1 Timeline of the experiment

Fig. 2 View of the concept mapper tool (translated from Russian)

Parallel versions of a test were used for pre- and post-tests. They both consisted of six open-ended questions and had a maximum score of twelve points. Questions aimed at checking students' understanding of the mechanisms and characteristics of

the process of state transitions, for example: “What happens with the energy of molecules during evaporation?”; “In what state(s) can H₂O be at the temperature of 0 °C at normal pressure?”; and “Do molecules change during a phase transition? Why?”.

Fig. 3 View of the online lab (English version, same as the Russian 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 View of the peer assessment tool for the criteria condition, with the first concept map (translated from Russian)

Giving feedback was done anonymously with the help of a special online peer assessment tool. Students could see the reviewed products and the assessment criteria (in the criteria condition) and use the tool to write their comments. The peer assessment tool used for both conditions is shown in Figs. 4 and 5.

Procedure

The experiment took place during regular school hours. The experimental lesson with the ILS was developed for one school lesson (45 min); students could work through the material at their own pace and return to previous phases in the ILS if needed. In the middle of the lesson and ten minutes before the end, students were informed

about the time and advised to move to the next phase(s) of the ILS; however, they could ignore this advice.

The activity of giving feedback was done at the end of the ILS. In the peer assessment tool, the researcher could check if all students had given feedback. Five minutes before the end of the lesson, those who had not yet given feedback were asked to do so. After giving feedback, students were asked to re-work their own concept maps. Their original concept map created in the Conceptualization phase could be loaded so that students could improve it. As with any other task in the inquiry process, they could choose to skip it.

The researcher was present during the lesson; participants could ask questions about the procedure or tools, but not about the content.

Fig. 5 View of the peer assessment tool for the free condition, with the second concept map (translated from Russian)

Give feedback on this concept map. Scroll down to write your comments.

Write your comments here.

Pre- and post-tests (about 15 min each) on the topic of state transitions were administered during the physics classes before and after the experimental lesson using the ILS, respectively. All three sessions took place within one week (secondary schools in Russia work six days), with several days between them.

Analysis

The pre- and post-test that were administered used open-ended questions that needed to be scored. The scoring was based on rubrics developed by the researchers and approved by participating teachers. The rubrics included the right answers and corresponding points. To check interrater reliability, a second rater graded 26% of the pre-tests and 17% of the post-tests. Cohen's kappa was 0.90 for pre-tests and 0.95 for post-tests, indicating almost perfect agreement.

Students' learning products generated during the lesson were also analyzed—their own concept maps and the feedback they gave. Together with post-test scores, learning products can give information about reviewers' learning, thus helping to answer the research questions.

The concept maps that were analyzed were the final version completed by students after finishing the main part of the ILS and after giving feedback. Concept maps were coded using the following criteria:

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

These criteria were based on the three-part scheme for coding concept maps suggested by Ruiz-Primo et al. (2001). Their first two criteria were used in the original way and the third one was replaced. Instead of the original convergence score (the proportion of correct links in the student's map out of all possible links in the criterion map), a complexity score was used. No criterion map was used, as different student concept maps could all be seen as valuable.

The coding was based on a general representation of the key concepts and connections between them, created jointly with an expert. All three scores were used separately for the analysis. The proposition accuracy score had no maximum, as students could use and link as many

relevant concepts as they wanted. The salience score had a maximum of one point. As mentioned above, the complexity scoring was developed by the researchers. It used a nominal scale and aimed at distinguishing between a linear construction (“sun” or “snake” shaped) and a more complex structure with more than one hierarchy level, assigning one point for a linear structure and two points for a more complex structure.

A second rater coded 15% of the concept maps with a very high level of agreement, as Cohen’s kappa was 0.82.

Finally, the quality of the feedback provided by students was assessed. In our study, high quality feedback meant correct and/or (potentially) useful comments made about a concept map. To be correct, a comment should be accurate in terms of physics content (for questions about concepts and mistakes, etc.) and in terms of the particular concept map (for questions about layout and examples). To be useful, a comment should be explained. For the question “Why is this concept map helpful or not helpful for understanding the topic?”, an example of a useful comment would be “it doesn’t help, as it has mistakes,” whereas an example of a useless comment would be “it doesn’t help, as I don’t understand it.”

Different coding schemes were used to operationalize this understanding of feedback quality for the two conditions, because of the different ways students gave feedback in each condition. For the free condition without provided criteria, the feedback was assessed in two successive steps:

- Step 1: focus of feedback—if a comment was content-related (“add gas”) and not emotion-related (“well done”), it received one point.

- Step 2: correct suggestions—each correct suggestion for a specific area of improvement received one point, each additional correct suggestion for the same area received half a point. The areas were expected to be similar to those addressed in the provided criteria: missing concepts, layout, mistakes, examples, and helpfulness.

For the condition with provided assessment criteria, each correct and useful answer received one point; similar to the other condition, any additional suggestion received half a point.

To get an overall score for the quality of feedback, the average of the two scores (one for each concept map) was used. Following the same procedure as with the other learning products, 17% of the feedback given (including both conditions) was coded by a second rater, with Cohen’s kappa being 0.90, which meant almost perfect agreement.

Results

The difference in pre-test scores between the conditions was not significant: $M_{\text{FREE}} = 3.41$ ($SD = 2.22$) and $M_{\text{CRITERIA}} = 3.31$ ($SD = 2.46$), $t(91) = 0.21$, $p = 0.83$ (see Table 2).

For further analysis, participants were divided into three groups based on their pre-test results: low prior knowledge (pre-test score greater than 1 SD below the mean), average prior knowledge (pre-test score within 1 SD above or below the mean), and high prior knowledge (pre-test score greater than 1 SD above the mean). The overall distribution of students between the low, average, and high prior knowledge groups in our sample was 23, 52, and 18 students, respectively.

Table 2 Test scores by knowledge level and condition (maximum score is 12)

	Low prior knowledge		Average prior knowledge		High prior knowledge		Total	
Free condition	(n = 9)		(n = 27)		(n = 8)		(n = 44)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre-test	0.56	0.53	3.33	1.18	6.88	0.84	3.41	2.22
Post-test	5.11	1.77	5.52	2.59	8.25	2.66	5.93	2.65
Criteria condition	(n = 14)		(n = 25)		(n = 10)		(n = 49)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre-test	0.64	0.50	3.32	1.18	7.00	1.41	3.31	2.46
Post-test	4.57	2.65	5.48	2.72	7.70	2.50	5.67	2.83
Total	(n = 23)		(n = 52)		(n = 18)		(n = 93)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre-test	0.61	0.50	3.33	1.17	6.94	1.16	3.35	2.34
Post-test	4.78	2.31	5.50	2.63	7.94	2.51	5.80	2.74

Learning Gain

First, we checked whether students learned during the experimental lesson. Paired t tests were used for pre- and post-test scores. The results showed that students learned during the lesson, both overall [$t(92) = 8.63$, $p < 0.01$, Cohen's $d = 0.89$] and in each condition [$t_{\text{FREE}}(43) = 6.74$, $p < 0.01$, Cohen's $d = 1.02$; $t_{\text{CRITERIA}}(48) = 5.61$, $p < 0.01$, Cohen's $d = 0.80$].

Descriptive statistics for all prior knowledge groups and both conditions are presented in Table 2.

The Effect of Condition and Prior Knowledge on Post-test Scores

An ANOVA was used to answer the research questions concerning whether being provided with assessment criteria when giving feedback influences the post-test scores, and whether this is different for different prior knowledge groups. Prior knowledge level and condition were used as independent variables and post-test score as the dependent variable. Only the main effect of prior knowledge was found to be significant [$F(2,87) = 8.24$, $p = 0.001$, $\eta^2 = 0.16$], meaning that students with higher prior knowledge had higher post-test results as well.

Learning Product Characteristics

A regression analysis was used with exploratory purposes to see if characteristics of the learning products (feedback quality and concept map quality) generated when working in the ILS explained the post-test scores. The descriptive statistics for these characteristics are given in Table 3. The coding schemes and maximum scores for these characteristics were discussed in the analysis section.

As feedback was given differently in the two conditions, the regression analyses were also conducted separately. To explore the influence of each variable, a backwards regression was used, with the following independent variables:

the quality of the feedback provided, the number of correct links, the proportion of correct links, the complexity of the concept map and the fact of changing the concept map after giving feedback. The fact of changing one's own concept map after giving feedback was added to the analysis, as not all students completed this task. Based on the significance of the regression, for the free condition only the quality of the feedback provided and the proportion of correct links were included in the analysis ($p = 0.047$), while for the criteria condition all variables were included ($p = 0.020$).

For both conditions, only the feedback quality was found to be a significant predictor of post-test scores [$F_{\text{FREE}}(2, 39) = 3.30$, $p < 0.05$, $F_{\text{CRITERIA}}(5, 43) = 3.04$, $p < 0.0$]. When the quality of feedback increased by one point, post-test scores increased by 1.45 points for the free condition and by 0.87 for the criteria condition. It is also worth mentioning that students in the free condition gave content-related feedback: the mean score for feedback quality is higher than one, which, according to the coding scheme, means that content-related feedback was given. As students from the other condition used provided assessment criteria and assessment criteria were quite straightforward, it was much easier for them to score at least one point for feedback quality.

The quality of students' own concept maps was compared between the conditions. None of the differences was statistically significant. However, it is important to mention that only 48% of students changed their own concept maps after giving feedback. This means that around half of the analyzed concept maps did not include any possible learning gain after viewing the two concept maps and giving feedback.

Conclusion and Discussion

Giving feedback is an important part of the peer-assessment process (e.g., Cho and MacArthur 2011; Ion et al. 2019). It consists of several steps, starting with defining the assessment criteria, and followed by assessing the product and providing suggestions for improvement (Sluijsmans 2002). As the first step in this process, defining assessment criteria can play a crucial role in the whole feedback-giving activity and the learning it generates. The aim of this study was to investigate the role that provided assessment criteria played in peer reviewers' learning.

To answer the research questions, two conditions were compared, giving feedback without or with provided assessment criteria, while taking into account students' prior knowledge. An ANOVA was conducted to find the effect of condition and prior knowledge on the post-test results. Based on previous findings (Gan and Hattie 2014; Orsmond et al. 2000; Panadero et al. 2013), an effect of condition and an interaction effect between condition and prior knowledge were expected. However, the findings did not reveal any significant differences,

Table 3 Mean scores for feedback and concept map (CM) quality by condition

Variable	Free condition ($n = 44$)		Criteria condition ($n = 49$)	
	M	SD	M	SD
Feedback quality	1.18	0.75	1.46	1.25
Number of correct links in the CM	4.10	2.98	3.14	2.10
Proportion of correct links in the CM	0.70	0.30	0.61	0.37
Complexity of the CM	1.55	0.50	1.41	0.61
Changed CM after giving feedback	0.57	0.50	0.41	0.50

neither between the conditions, nor between conditions in combination with different levels of prior knowledge (interaction effect). Quite as expected, the main effect of prior knowledge was found to be significant, meaning that students with higher pre-test scores also had higher post-test scores. It was also verified that students learned during the lesson, as their test scores improved significantly from pre- to post-test, both as a whole group and in each condition. This can be viewed as a prerequisite for further investigation of the learning process, including the influence of a feedback-giving activity.

A plausible explanation for the lack of an effect of condition could be that the peer assessment task was too challenging for students for this domain and topic. Könings et al. (2019) found that guidance in the peer assessment process, such as provided assessment criteria, led to lower domain-specific accuracy on the post-test. The authors explained such a surprising result by the fact that students apparently did not have enough domain knowledge, and could have been overwhelmed by the extra mental load caused by the peer assessment tasks. This conclusion was based on the rather low post-test scores of their participants. In our study, even though students did learn, they did not learn a lot, with mean post-test scores at 5.8 out of 12. Moreover, they had to complete the new challenging task of creating a concept map for the learning material. Taken together, this could mean that they were not fully ready to give feedback. This extra difficulty could have led to less of a difference between the conditions and, as a result, to the lack of an intervention effect.

Such task complexity could have also been connected with the product students had to assess. Concept mapping in general is a skill that might require some training. During our experiment, students were exposed to at least two new tasks—concept mapping and giving feedback on a concept map—in the same lesson. This meant that in a relatively short time, they had to perform several new activities with new domain material. This could have been too challenging for students. The fact that around 15% of participants did not complete the feedback-giving task pointed in this direction. To confirm or disconfirm this line of reasoning, new research that includes training in giving feedback and creating concept maps could be conducted. Some training or more experience with giving feedback have been shown to be a good way to develop such a skill (e.g., Rotsaert et al. 2018; Sluijsmans et al. 2002); therefore, a training activity before the intervention could decrease the complexity and thereby make the differences between the conditions more obvious.

The quality of the feedback given by students was assessed as well. As the way feedback was given differed between the conditions, it was not possible to compare these scores statistically. A separate regression analysis was done for each condition, which showed that the quality of the feedback given was a significant predictor of post-test

scores for both conditions. However, it is difficult to draw any conclusions based on this result, as both higher feedback quality and higher post-test scores could result from higher prior knowledge. Nevertheless, this result can be a direction for practical implications. Encouraging students to invest more effort in giving feedback may lead to higher quality feedback and, as a result, to more domain learning.

Another aspect that is worth mentioning is that students in the free condition gave content-related feedback. This probably implies that they were able to think about their own assessment criteria and gave feedback based on those criteria. This result is aligned with the findings by Tsivitanidou et al. (2011), who showed that students had basic skills for peer assessment and could do it to some extent without being supported by provided assessment criteria.

The obtained results did not give a straightforward answer to the main research question: what role being provided with assessment criteria plays in reviewers' learning. The main conclusions are as follows: first, post-test scores of the students in the free condition were not lower than in the criteria condition, and the quality of students' own concept maps was also equal in both conditions. Second, students in the free condition were able to give content-related feedback even without being provided with assessment criteria. In terms of implementation, this suggests that giving students freedom to create their own assessment criteria will not lead to less learning. We even observed trends pointing in the direction that there might be some differences between the two conditions in favor of the free condition. The take-away message is that even without being provided with assessment criteria, secondary students are able to give content-related feedback.

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. 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.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Statement Carrying out of the research was approved by the Ethical Committee of the University.

Consent Statement An appropriate consent form was used for the participants.

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

- Canty, D., Seery, N., Hartell, E., & Doyle, A. (2017, July 10–14). *Integrating peer assessment in technology education through adaptive comparative judgment*. Paper presented at the PATT34, Millersville University, PA, USA.
- Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology*, 103(1), 73–84. <https://doi.org/10.1037/a0021950>
- de Jong, T. (2006). Technological advances in inquiry learning. *Science*, 312, 532–533.
- de Jong, T. (2019). Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning*, 35(2), 153–167. <https://doi.org/10.1111/jcal.12337>
- Deiglmayr, A. (2018). Instructional scaffolds for learning from formative peer assessment: effects of core task, peer feedback, and dialogue. *European Journal of Psychology of Education*, 33(1), 185–198. <https://doi.org/10.1007/s10212-017-0355-8>
- Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology*, 21(1), 49–58. [https://doi.org/10.1016/S0193-3973\(99\)00050-7](https://doi.org/10.1016/S0193-3973(99)00050-7)
- Falchikov, N. (2004). Involving students in assessment. *Psychology Learning & Teaching*, 3(2), 102–108. <https://doi.org/10.2304/plat.2003.3.2.102>
- Friesen, S., & Scott, D. (2013). *Inquiry-based learning: A review of the research literature*. Paper prepared for the Alberta Ministry of Education.
- Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 42(6), 861–878. <https://doi.org/10.1007/s11251-014-9319-4>
- Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, 52, 315–325. <https://doi.org/10.1016/j.chb.2015.06.019>
- Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, 44(1), 124–138. <https://doi.org/10.1080/02602938.2018.1484881>
- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787. <https://doi.org/10.1080/03075079.2013.821974>
- Könings, K. D., van Zundert, M., & van Merriënboer, J. J. G. (2019). Scaffolding peer-assessment skills: Risk of interference with learning domain-specific skills? *Learning and Instruction*, 60, 85–94. <https://doi.org/10.1016/j.learninstruc.2018.11.007>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193–211. <https://doi.org/10.1080/02602938.2019.1620679>
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496. <https://doi.org/10.1002/tea.20347>
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23–38. <https://doi.org/10.1080/02602930050025006>
- Panadero, E., Romero, M., & Srijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195–203. <https://doi.org/10.1016/j.stueduc.2013.10.005>
- Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science*, 43(5), 591–614. <https://doi.org/10.1007/s11251-015-9353-x>
- Pedaste, M., Mäeots, 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>
- Peters, O., Körndle, H., & Narciss, S. (2018). Effects of a formative assessment script on how vocational students generate formative feedback to a peer's or their own performance. *European Journal of Psychology of Education*, 33(1), 117–143. <https://doi.org/10.1007/s10212-017-0344-y>
- Rosa, S. S., Coutinho, C. P., & Flores, M. A. (2016). Online peer assessment: method and digital technologies. *Procedia-Social and Behavioral Sciences*, 228, 418–423.
- Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2018). “Now you know what you're doing right and wrong!” Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*, 33(2), 255–275. <https://doi.org/10.1007/s10212-017-0329-x>
- Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching*, 38(2), 260–278. [https://doi.org/10.1002/1098-2736\(200102\)38:2%3c260::AID-TEA1005%3e3.0.CO;2-F](https://doi.org/10.1002/1098-2736(200102)38:2%3c260::AID-TEA1005%3e3.0.CO;2-F)
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment. The training of peer assessment skills* [unpublished doctoral dissertation]. Open University of the Netherlands, The Netherlands.
- Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education*, 27(5), 443–454. <https://doi.org/10.1080/0260293022000009311>
- Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21(4), 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(1), 194–204. <https://doi.org/10.1080/10494820.2013.834828>

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