



Learning from reviewing peers' concept maps in an inquiry context: Commenting or grading, which is better?

Natasha Dmoshinskaia^{*}, Hannie Gijlers, Ton de Jong

Department of Instructional Technology, Faculty of Behavioral, Management, and Social Sciences (BMS) University of Twente, Enschede, the Netherlands

ARTICLE INFO

Keywords:

Student evaluation
Evaluation methods
Peer assessment
Giving feedback

ABSTRACT

In peer assessment, both receiving feedback and giving feedback (reviewing peers' products) have been found to be beneficial for learning. However, the different ways to give feedback and their influence on learning have not been studied enough. This experimental study compared giving feedback by writing comments and by grading, to determine which contributes more to the feedback providers' learning. Secondary school students from Russia ($n = 51$) and the Netherlands ($n = 42$) gave feedback on concept maps during a physics lesson. The lesson was given in an online inquiry learning environment that included an online lab. Students gave feedback in a special Peer Assessment tool, which also provided assessment criteria. Findings indicate that post-test knowledge scores were higher for students from the commenting group. The difference between the groups was largest for the low prior knowledge students. Possible educational implications and directions for further research are discussed.

1. Introduction

Peer assessment is becoming more and more popular among educators. According to a meta-analysis conducted by Li, Xiong, Hunter, Guo, and Tywoniw (2020), peer assessment has a positive, average-sized effect on students' learning. The results of the analyzed studies also indicated that students develop reflection and (self-) evaluation skills through peer assessment and they feel more responsibility for their learning. Moreover, the same meta-analysis showed that computer-based peer assessment leads to bigger learning gains than paper-based peer assessment. Another meta-analysis of technology-facilitated peer assessment (Zheng, Zhang, & Cui, 2020) also showed that this type of peer assessment has a positive effect on learning compared to paper-based peer feedback, with an overall mean effect size of 0.54. Despite the ongoing research in this area, it is not yet fully clear how different characteristics of the peer assessment process influence its presumed effect. Investigating these issues by focusing on particular aspects and mechanisms of the peer assessment process with a (quasi-) experimental design can especially contribute to knowledge about this process (Strijbos & Sluijsmans, 2010).

Peer assessment has two components – giving feedback to and receiving feedback from peers. The definition of feedback used in the current study is based on the work of Hattie and Timperley (2007), who viewed it “as information provided by an agent (e.g., teacher, peer,

book, parent, self, experience) regarding aspects of one's performance or understanding” (p. 81). According to the same authors, effective feedback should cover three main questions: Where am I going? How am I going? and Where to next? The first question is associated with the desired state, the second question indicates the progress so far, and the third one suggests the next step.

The majority of studies have focused on one part of the peer assessment process, namely, receiving peer feedback (Cao, Yu, & Huang, 2019). The explanation for that can be that receiving feedback is often regarded as very beneficial for students. The reasons for such benefits come from the fact that receiving peer feedback gives learners additional and more varied feedback compared to feedback only from their teacher, and this extra feedback may help them to improve their performance (Cho & MacArthur, 2010; Falchikov, 2013; Li, Xiong, Hunter, Guo, & Tywoniw, 2020; Topping, 1998). For example, receiving peers' feedback can lead to a higher score on an exam or better quality of a student-created learning product such as an essay, a poster, or a webpage.

The other part of a peer-assessment activity – giving feedback to peers or, in other words, reviewing – is much less studied. However, a few studies (Li, Liu, & Steckelberg, 2010; Lundstrom & Baker, 2009; Phillips, 2016) have shown not only that students learn from giving feedback, but also that they may learn even more from giving than from receiving it. This can be explained by the fact that students who give

^{*} Corresponding author at: P.O. Box 217, 7500 AE Enschede, The Netherlands.

E-mail addresses: n.dmoshinskaia@utwente.nl (N. Dmoshinskaia), a.h.gijlers@utwente.nl (H. Gijlers), a.j.m.dejong@utwente.nl (T. de Jong).

feedback (the reviewers) must perform cognitive activity to evaluate their peers' products, which would include thinking of assessment criteria, comparing a piece of work with the required state, and providing suggestions for improvement. According to [van Popta, Kral, Camp, Martens, and Simons \(2017\)](#), giving feedback should be seen as a learning activity. Their literature review concluded that a student-reviewer benefits in terms of the following activities and outcomes: higher-level thinking, critical reflection and insight, improving their own product, meaning making and knowledge building, and the ability to develop evaluative judgements.

1.1. Giving feedback

Giving feedback to peers consists of several steps. Studies on this topic have been conducted over several decades, in various contexts ([Cho & Cho, 2011](#); [Flower, Hayes, Carey, Schriver, & Stratman, 1986](#); [Hayes, Flower, Schriver, Stratman, & Carey, 1987](#); [Patchan & Schunn, 2015](#); [Sluijsmans, 2002](#)) and the steps identified by different authors are rather similar. For example, [Sluijsmans \(2002\)](#) suggested a model for giving peer feedback that consists of three main steps:

- define assessment criteria,
- judge the performance of a peer,
- provide feedback for future learning.

This view on the feedback-giving process is supported by a study by [Cho and Cho \(2011\)](#), who investigated learning achieved by reviewers through giving feedback on peers' technical reports. They concluded that "reviewers learn by explaining what makes peer texts good or bad, by identifying problems that exist in those peer texts, and then in devising ways in which those problems can be solved" (p. 630).

The present study focuses on investigating the second and third steps of the model for giving feedback suggested by [Sluijsmans \(2002\)](#) – judging performance and providing directions, with the assessment criteria being provided to students.

The type of feedback is an important factor determining the effect of giving feedback on the reviewer's learning process. A study by [Lu and Zhang \(2012\)](#) compared the influence of providing different types of feedback – affective and cognitive – on learners' performance. While affective feedback operates mostly with evaluative statements – positive (praise) or negative (critique), cognitive feedback focuses on the nature of the task itself. The results of this study demonstrated that only the giving of cognitive feedback contributed to the reviewers' learning outcomes; several studies described below have emphasized that providing suggestions on how peers can improve their products or performance is very important for the reviewer's learning gains. [Wooley, Was, Schunn, and Dalton \(2008\)](#) analyzed the impact of the type of feedback on reviewers' performance. They found that the group of university students who had to give elaborated comments together with a grade had higher quality writing than those who gave only grades. According to these authors, this could be attributed to the fact that reviewers were more cognitively involved when elaborating than when grading. They argued that there was a strong connection between articulating and thinking. The need to provide a detailed comment led to deeper thinking about the material, which not only facilitated evaluation of peers' work, but also triggered reflection on the reviewers' own writing. These findings are supported by work by [Xiao and Lucking \(2008\)](#), who studied over 200 undergraduates and found that giving feedback by providing comments and suggestions together with a grade led to higher quality of reviewers' own writing than giving feedback by providing a grade only. This was corroborated in more detail in a study by [Lu and Law \(2012\)](#), who studied secondary school children who reviewed their peers' school projects. These authors found that the number of problems identified and improvements suggested correlated positively with reviewers' performance. According to the authors, this can be explained by the fact that spotting mistakes and coming up with

solutions activate cognitive processes critical for reviewers' learning. Based on these findings, it seems likely that one way to facilitate learning from the feedback-giving process is to encourage students to give meaningful comments and not just grades for peers' products.

In the literature, two additional factors have been identified that mediate the effect on the reviewer's learning of giving feedback to peers: the quality of the products that are being reviewed and students' prior knowledge.

Diversity in the reviewed products may facilitate learning. Studies have indicated that both commenting on positive features of a product ([Cho & Cho, 2011](#)) and providing critical feedback to peers can contribute to a reviewer's learning ([Cho & Cho, 2011](#); [Li et al., 2010](#); [Lu & Zhang, 2012](#)). When inspecting and reviewing good examples, students see successful strategies at work, and can adopt new strategies or verify known ones. By reviewing lower-level pieces of work, they can practice such skills as diagnosing and detecting problems, as well as suggesting solutions. However, the learning of peer reviewers can be hindered when the reviewed products are of too low quality. In the study conducted by [Alqassab, Strijbos, and Ufer \(2018\)](#), students gave feedback on geometry proofs that differed in quality: either almost correct or full of mistakes. Participants reviewing almost correct proofs demonstrated better understanding of the topic and provided more accurate feedback than those reviewing proofs with errors. To balance the effect of high and low quality of the reviewed products in the current study, the quality of reviewed products was controlled by offering all students the same set of lower and higher quality products.

Students' prior knowledge has been shown to influence their learning from giving feedback; in order to learn from giving feedback, students should have enough domain knowledge to be able to give correct and meaningful feedback. In a study conducted by [van Zundert, Könings, Sluijsmans, and van Merriënboer \(2012\)](#), secondary school students were divided into two groups: students in the first group were instructed about a new domain (how to perform scientific investigations) before reviewing peers' performance on this same task, whereas the other group had to give feedback while being instructed at the same time. Students from the first group showed higher improvement in both domain knowledge and performance in giving feedback than students from the second group. These findings are in line with the outcomes of a study by [Alqassab, Strijbos, and Ufer \(2018\)](#), who found that low prior knowledge students could provide feedback only about the correctness of the product, whereas higher prior knowledge students could also comment on a conceptual level, triggering reflection about the task and learning goals. To sum up, reviewers' prior knowledge influences how well they perform a feedback-giving activity and thereby the learning it can engender.

Moreover, the combination of reviewers' prior knowledge and different levels of quality of the reviewed products can create an interaction effect. In a study by [Patchan \(2011\)](#), highly skilled writers benefited equally from reviewing texts with different levels of quality, whereas less skilled writers benefited more from reviewing texts of lower quality. This result is supported by other research; for example, [van Zundert, Sluijsmans, Könings, and van Merriënboer \(2012\)](#) discovered that increasing the complexity of the reviewing task may lead to cognitive overload, resulting in poor performance in giving feedback. As the complexity of the same task can be perceived differently by students with different prior knowledge levels, prior knowledge should be taken into account when investigating the feedback-giving process.

1.2. Research questions

Several studies ([Lu & Law, 2012](#); [Wooley, Was, Schunn, & Dalton, 2008](#); [Xiao & Lucking, 2008](#)) have indicated that giving comments as a part of the feedback is more beneficial for reviewers' own knowledge development than just grading peers' work. However, these studies covered rather extensive products, such as a piece of writing or a six-week school project. In addition, the work done thus far has focused

primarily on university-level students. The finding that commenting contributes to reviewers' learning more than grading may not reflect the situation in secondary school or may not be true for smaller scale learning products that require less time and effort to be invested by the reviewer.

In the current study, we further investigate the effects of giving feedback on reviewers' learning, but now in the context of secondary education. In doing so, our focus is on smaller scale learning products (i. e., concept maps, rather than pieces of writing or extended projects), since this fits better with this age group and with the STEM (science, technology, engineering and math) domains we are investigating. Moreover, as shown in several studies (Alqassab et al., 2018b; Patchan, 2011; van Zundert, Könings et al., 2012) learning from giving feedback can be different for students with different prior knowledge levels when being asked to give feedback on products with diverse levels of quality. Therefore, investigating the effect of prior knowledge on reviewers' learning can have practical implications.

Thus, the aim of the study was to investigate which form of feedback being given – comments or grades – contributes more to reviewers' learning in a secondary school STEM context and whether this contribution was different for students with different levels of prior knowledge. Learning was broadly construed, and was measured via several indicators: domain knowledge tests, the reviewers' own learning products and the quality of the provided feedback. Prior knowledge groups (low, average, and high) were used for practical reasons; if learning does differ for different prior knowledge groups when giving these two forms of feedback, having such groups identified would make practical implications for the classroom easier: (potentially) different recommendations for different prior knowledge groups. The main research question is formulated as follows: Which way of providing feedback is more beneficial for a peer reviewer: commenting or grading? There is also a secondary research question: Is there a differential effect for students with different prior knowledge levels?

2. Method

2.1. Participants

The data set initially consisted of 139 participants representing two countries – Russia ($n = 81$) and the Netherlands ($n = 58$), with $M_{AGE} = 14.55$ years old ($SD = 0.49$). In Russia, students came from three eighth grade classes of a comprehensive secondary school, while in the Netherlands they were from a bilingual pre-university educational track. The only exclusion criterion used was absence from part of the study, which reduced the total number of participants to 93 (42 boys and 51 girls): 51 from Russia, and 42 from the Netherlands. The distribution between the conditions was nearly equal: commenting – 46 and grading – 47.

Eighth grade was chosen based on convenience sampling – the researchers were looking for a topic that would be addressed in a secondary school STEM context in both countries, and found an appropriate topic in the eighth-grade curriculum. The two countries are those where the researchers have contacts and access to students. Though students represented two different countries, they were very similar in key aspects: they had no experience working with online inquiry learning environments; they were the same age ($M_R = 14.64$, $SD = 0.36$; $M_{NL} = 14.45$, $SD = 0.60$), and their pre-test scores did not show a statistically significant difference [$M_R = 3.77$, $SD = 2.23$; $M_{NL} = 3.77$, $SD = 2.30$; $t(91) = -0.02$, $p = .99$]. Moreover, even though their teachers reported that students in both countries were familiar with the idea of peer assessment, the students did not have any experience with giving feedback in online inquiry learning environments, nor did they receive any specific training in doing this. Therefore, both groups were analyzed together.

To eliminate any possible differences between schools and classes, participants were randomly assigned to one of the two experimental

conditions in each class. The conditions involved giving feedback by providing comments or giving feedback by grading the product with one of five smileys (a range of faces, going from a very unhappy face to a very happy one).

2.2. Study design

This is an experimental study, using a two-group pre-test post-test design, in which students had to give feedback on two concept maps. Participants in both conditions were supported in doing this by being given assessment criteria. These criteria were based on the above-mentioned three-step approach to giving feedback described by Sluijsmans (2002) and Hattie and Timperley (2007). These criteria introduced important characteristics of a concept map (missing concepts, structure, links, etc.), which were based on the criteria described in the study by van Dijk and Lazonder (2013). The assessment criteria, although following similar principles, were worded differently for the two conditions, as can be seen in Table 1. In the comment condition, students had to answer the open-ended questions by typing their comments, and in the smiley condition, students had to answer the questions by choosing a relevant smiley.

All participants received the same concept maps to give feedback on. These concept maps were constructed as if they came from peers; students were told that these concept maps came from some students who were not necessarily from their class. They were asked to give feedback with a formative and not a summative purpose; moreover, they were encouraged to provide constructive critical feedback to improve these peers' work. Several studies (e.g., Patchan, 2011; Patchan & Schunn, 2015) have shown that the quality of the reviewed work influences the quality of the provided feedback and the learning gains. In our study, to create equal conditions and eliminate possible differences in the learning products to be reviewed, all students reviewed the same set – one good quality and one poor quality concept map.

2.3. Materials

The concept maps that students evaluated were presented in an online inquiry learning space (ILS). ILSs are created with the Go-Lab ecosystem (see www.golabz.eu and de Jong, Sotiriou, & Gillet, 2014). An ILS follows the principles of inquiry learning: students perform investigations with an online laboratory and follow the different stages of an inquiry cycle (Pedaste et al., 2015). Go-Lab ILSs also provide students with tools that scaffold inquiry processes (such as a scratchpad that supports the creation of hypotheses) and include all types of multimedia material in the different stages of inquiry.

The ILS that was used for the experiment was about the physics topic of convection. This topic is part of the heat transfer theme in the curriculum in both countries. During the lesson, students could work through the ILS at their own pace and return to previous stages if necessary. The ILS included the following stages:

- Orientation – The topic was introduced by a short video and the research question was set. The question was formulated as a real-life

Table 1
Assessment criteria for giving feedback.

Comment condition	Smiley condition
Which concepts are missing?	Does the concept map include all core concepts for the topic?
How would you change the structure of the map?	Does the concept map have a good structure?
Which links should be renamed to be more meaningful?	Are all the links meaningful?
Which examples should be added?	Are there any examples?
Why does or does not the concept map help with understanding the topic?	Does the concept map help you with understanding the topic?

situation, which should trigger students' inquiry process. The question was: Would we feel equally warm sitting on a sofa in a room with a low and a high ceiling when the heating system is on?

- **Conceptualization** – The stages of scientific experimentation were mentioned to students. They were asked to create a concept map about convection to demonstrate their ideas about the topic, which was done with a help of a Concept Mapper tool (see Fig. 1). The concept map included pre-defined terms and names of links, as well as an opportunity to add new terms and rename links. Pre-defined concepts and links were used as scaffolds in the process of creating a concept map, as they gave students a starting point.
- **Investigation** – Students were asked to formulate their hypotheses, and could then check them in an online lab. To scaffold students' experimentation, a hypothesis scratchpad was used. This tool included pre-defined terms and half of a hypothesis to direct students in their investigation. The lab allowed changing the height of the ceiling and checking the temperature at different heights; see Fig. 2.
- **Conclusion** – Students tried to answer the research question based on the observations they had made using the online lab.
- **Discussion** – Students gave feedback on two concept maps and had an opportunity to improve their own concept map if desired.

Students were asked to assess two concept maps. One was very low in quality and included only a few concepts. The other had many more concepts and better-named relationships between them, but did not contain examples. However, this concept map also included a common misconception, that is, that convection can occur in solids. The concept maps were presented to all participants in the same order: lower quality first, higher quality second. This was done so the students did not use examples from the higher quality concept map as suggestions for improving the lower quality one.

Giving feedback was done in the special peer assessment tool. This tool showed the product to give feedback on and the rubrics that guided students through this process. As an example, the higher quality concept map with the rubric for the grading (smileys) condition is shown in Fig. 3.

Pre- and post-tests were used, covering the same testing material. The test consisted of six open-ended questions and had a maximum possible score of 10 points; the number of points per question varied

from 1 to 3. It checked students' knowledge by asking them to explain topic-related concepts and phenomena or to apply theoretical knowledge to practical cases. Open-ended questions were chosen because giving feedback in general, and giving feedback about a concept map in particular, contribute to deeper understanding of the ideas and connections between them. Using the terminology of the revised Bloom's taxonomy (Krathwohl, 2002), our assessments consisted of questions checking not just remembering, but also understanding, applying, and analyzing the material.

The students' answers were graded by the researcher, with the score depending on the correctness of the answer and the level of reasoning displayed (see Table 2 for an example).

2.4. Procedure

In both countries, the study took place as part of the regular school lesson and was conducted in the same language as the teaching of physics: Russian in Russia and English in the Netherlands (as participating classes followed a bilingual program). During the experimental lesson, students were instructed to work individually and independently in the ILS and to follow the stages and instructions there, which included giving feedback on two concept maps. The ILS was intended to take up one school hour (50 min) and students could decide for themselves how to divide the time between the different stages of the ILS. The researcher indicated the amount of time left for students in the middle of the lesson and five minutes before the end. The researcher was present during the whole lesson; students could ask questions about the environment or the procedure, but not about the content.

Giving feedback was done anonymously through the peer assessment tool in the learning environment. In the tool, the researcher could see which students had given their feedback. Five minutes before the end of the lesson, students who had not yet given their feedback were asked to do so. All participants whose data were analyzed gave their feedback during the lesson.

After giving feedback, students were encouraged to improve their own initial concept maps, but it was not obligatory.

Pre- and post-tests (10–15 min) with the same test material were administered twice, once within a week prior to working in the ILS and once within a week afterwards. In both countries, this was done the

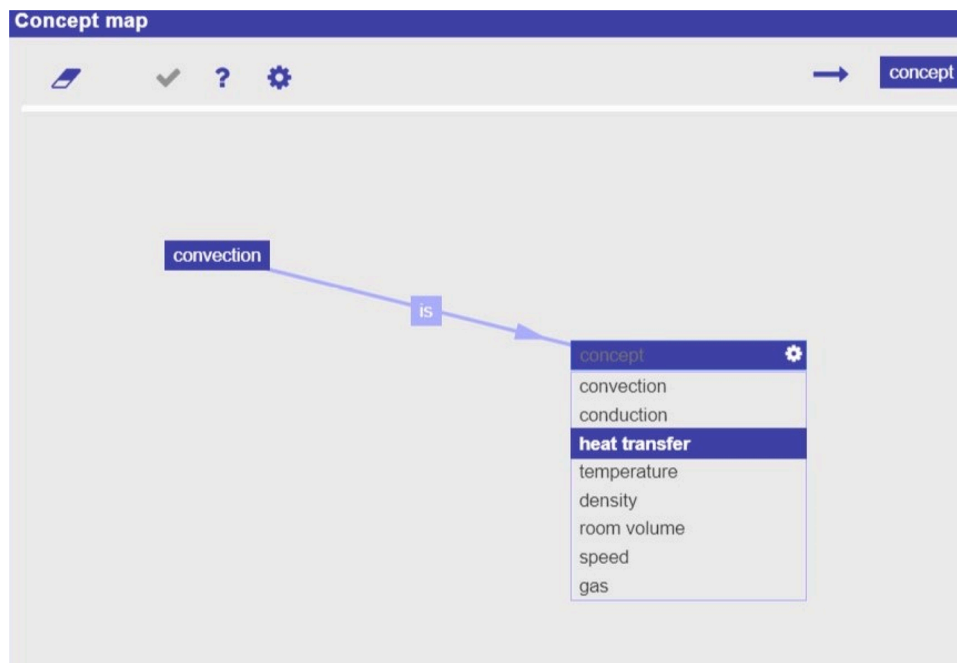


Fig. 1. View of the Concept Mapper tool (translated).

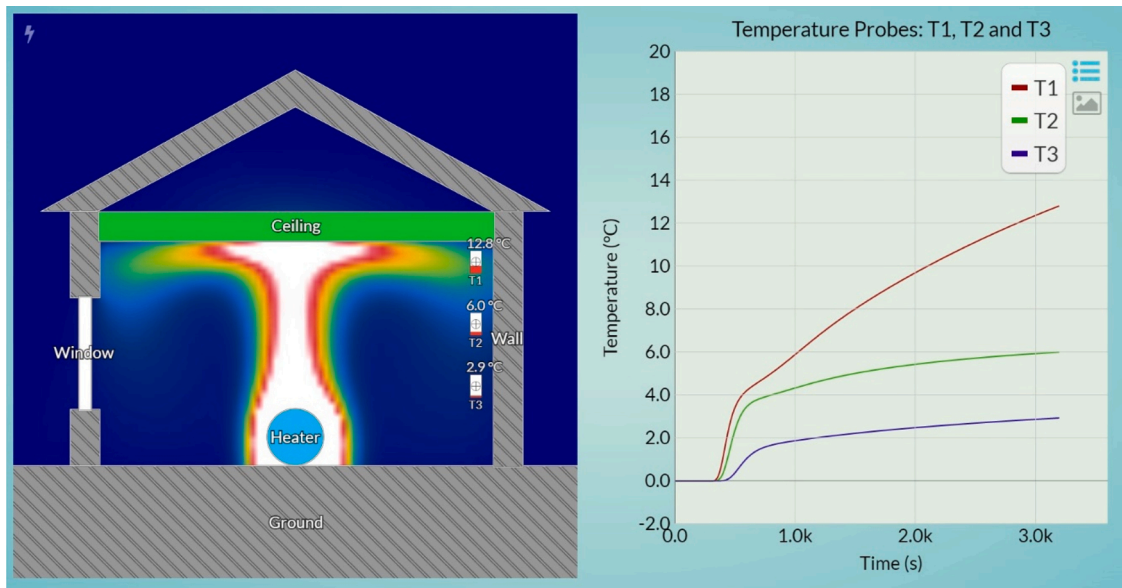


Fig. 2. View of the online lab “Vertical temperature gradients”. Images by The Concord Consortium, licensed under CC-BY 4.0 <https://concord.org/>.



Fig. 3. View of the peer assessment tool with a higher quality concept map, for the smiley condition (translated).

usual way other tests are done; in Russia it was a pencil-and-paper test and in the Netherlands it was done on the computer.

2.5. Analysis

Since the aim was to find out whether different ways of giving feedback (conditions) and different levels of prior knowledge influence learning, pre- and post-test scores were analyzed. To check the interrater reliability, 10 % of the knowledge tests were graded by a second rater.

Cohen’s kappa was .82 for Russia and .88 for the Netherlands.

Additionally, the quality of the final version of the students’ concept map (after giving feedback) was scored by coding it according to the following scheme:

- Proposition accuracy score – the number of correct links
- Salience score – the proportion of correct links out of total links
- Complexity score – hierarchy level of a concept map.

Table 2
Example of the grading scheme for test answers.

Question	Answers	Points
In a room there are two identical plants hanging on the wall. One is at a height of 50 cm, the other is at a height of 150 cm. Do you need to water them the same amount? Why?	No, differently.	1
	No, differently. The upper plant would need more water.	2
	No, differently. The upper plant would need more water because the temperature is higher in the upper part of the room so the water evaporates faster.	3

This scheme is based on the study by Ruiz-Primo, Schultz, Li, and Shavelson (2001), where concept maps were evaluated for accuracy and comprehensiveness. In that work, each student received three types of scores: proposition accuracy score – the number of correct propositions; convergence score – the proportion of accurate propositions in a student’s map out of all possible propositions in the criterion map; and salience score – the proportion of valid propositions out of all propositions in the student’s map.

In our study, students could include not only pre-defined terms, but also their own in the concept map, so their concept maps could differ from each other and from the expert map. For this reason, a complexity score was used instead of the convergence score. The complexity score had a scale from one to three, with one meaning only linear connections with no layers, two meaning a multilevel map and three meaning a multilevel map with cross connections. Ten percent of the concept maps were graded by a second rater, with adequate interrater reliability; Cohen’s kappa for Russia was .67 and for the Netherlands it was .73.

Finally, the quality of the feedback given by the students was assessed by coding the feedback that was provided. For the smileys condition, one point was given for each correct evaluation. A correct evaluation included either of two smileys with a similar meaning (for example, a happy and a very happy face) to avoid discrepancy in understanding smileys. As no specific training was done for students to assign each smiley with a particular value, our main goal was to see if students reacted to mistakes or incompleteness, as well as if they distinguished between the concept maps of different levels. In other words, for a low-quality concept map, both a very unhappy and an unhappy face would be a correct evaluation for the question about including all key concepts, while a happy or a very happy face would, in this case, be an incorrect evaluation. For the comment condition, one point was given for a meaningful suggestion/comment. The student’s score was an average of the scores received for assessing each concept map. To check the interrater reliability, 10 % of the feedback was assessed by a second rater; Cohen’s kappa was .78 for Russia and .80 for the Netherlands.

The characteristics of students’ own concept maps and the feedback they provided were used in an exploratory analysis of their connection with the post-test results.

3. Results

First, the distribution of prior knowledge between conditions was compared to check for inequality. The difference between pre-test scores for the two conditions was not statistically significant [$t(91) = 0.96, p = .34$].

Based on the pre-test results, students were divided into three groups: low prior knowledge (pre-test score more than 1 SD below the mean; 15 students), average prior knowledge (pre-test score within 1 SD above or below the mean; 58 students), and high prior knowledge (pre-test score more than 1 SD above the mean; 20 students). Division into prior knowledge groups was done only for the purpose of our analysis and had no bearing on the random assignment of students to one of the conditions.

Pre- and post-test scores were used for the analysis. The descriptive statistics for participants’ test scores per prior knowledge level and

condition are presented in Table 3.

Second, a normality check for the post-test scores was conducted for both conditions and the prior knowledge groups. A Shapiro-Wilk’s test and a visual inspection of the graphs showed that the post-test scores were approximately normally distributed for both conditions ($p_{COMMENT} = .43, p_{SMILEY} = .50$), with skewness of 0.256 ($SE = 0.350$) and kurtosis of $-0.575 (SE = 0.688)$ for the comment condition, and skewness of 0.107 ($SE = 0.347$) and kurtosis of $-0.482 (SE = 0.681)$ for the smiley condition. A Shapiro-Wilk’s test and a visual inspection of the graphs showed that the post-test scores were approximately normally distributed for all prior knowledge groups ($p_{LOW} = .32, p_{AVERAGE} = .13, p_{HIGH} = .48$), with skewness of 0.597 ($SE = 0.580$) and kurtosis of $-0.350 (SE = 1.121)$ for the low prior knowledge group, skewness of 0.535 ($SE = 0.314$) and kurtosis of 0.561 ($SE = 0.618$) for the average prior knowledge group, and skewness of $-0.339 (SE = 0.512)$ and kurtosis of $-0.089 (SE = 0.992)$ for the high prior knowledge group.

Third, an ANOVA was conducted with post-test score as the dependent variable, and condition and prior knowledge level as the independent variables, to answer the research question about the effect of different ways of giving feedback on learning.

Both main effects were found to be significant: prior knowledge level [$F(2,87) = 19.32, p < .001, \eta^2 = .31$] and condition [$F(1, 87) = 5.84, p = .018, \eta^2 = .06$]. The post-test scores of students from different prior knowledge groups were significantly different from each other ($p < .001$), with the estimated scores $M_{LOW} = 3.03 (SD = 0.50), M_{AVERAGE} = 4.13 (SD = 0.24), M_{HIGH} = 6.82 (SD = 0.44)$, indicating that having a higher pre-test score would lead to a higher post-test score. The post-test scores of students in the commenting condition were significantly higher ($p = .018$) than those for the grading (smiley) condition, with estimated average scores $M_{COMMENT} = 5.23 (SD = 0.33)$ and $M_{SMILEY} = 4.09 (SD = 0.34)$.

Apart from the main effects, a significant interaction effect between condition and prior knowledge level was also found, $F(2, 87) = 4.19, p = .018, \eta^2 = .09$ (Fig. 4). This means that conditions worked differently for students with different levels of prior knowledge.

To further specify this interaction effect, a separate analysis was conducted for each prior knowledge group. As the low and high ability

Table 3
Mean (SD) test scores (max score of 10) by prior knowledge level and condition.

	Low prior knowledge <i>n</i> = 10	Average prior knowledge <i>n</i> = 30	High prior knowledge <i>n</i> = 6	Total <i>n</i> = 46
Comment condition				
Pre-test	0.65 (0.47)	3.75 (1.30)	7.33 (1.47)	3.54 (2.28)
Post-test	4.45 (2.42)	3.98 (1.69)	7.25 (1.26)	4.51 (2.09)
Smiley condition				
	<i>n</i> = 5	<i>n</i> = 28	<i>n</i> = 14	<i>n</i> = 47
Pre-test	0.90 (0.22)	3.11 (1.20)	6.86 (0.57)	3.99 (2.23)
Post-test	1.60 (1.25)	4.27 (1.99)	6.39 (1.53)	4.62 (2.27)
Total				
	<i>n</i> = 15	<i>n</i> = 58	<i>n</i> = 20	<i>n</i> = 93
Pre-test	0.73 (0.42)	3.44 (1.28)	7.00 (0.92)	3.77 (2.25)
Post-test	3.50 (2.48)	4.12 (1.83)	6.65 (1.48)	4.57 (2.17)

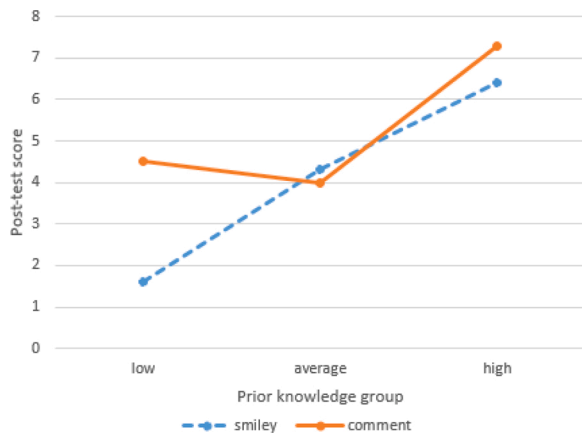


Fig. 4. Post-test scores per prior knowledge group.

groups were small, nonparametric tests were used to see the difference between the conditions for each group. As this analysis was post hoc for the same dataset as for the ANOVA, a Bonferroni correction was applied, leading to a statistical significance cutoff value of .025 instead of .05. The difference between conditions was not statistically significant for any of the prior knowledge groups [$H_{LOW}(1) = 4.91, p = .027, H_{AVERAGE}(1) = 0.22, p = .639, H_{HIGH}(1) = 1.26, p = .261$].

Finally, an exploratory regression analysis was conducted to obtain better understanding of the relation between the students' learning process and their learning performance. To do so, the analysis included the characteristics of artefacts produced by students (concept maps and feedback) as predictors and post-test scores as outcomes. The following variables were used to predict the post-test score for each student: the fact of changing their own concept map after giving feedback, quality of the final concept map as described above (proposition accuracy score, salience score, and complexity score) and quality of feedback given. The descriptive statistics for these variables are presented in Table 4. The complete overview of the coding procedure was given in the analysis section.

As the model was built for exploratory purposes, a stepwise backwards regression was used. The best fit was found for the model with the salience of the concept map excluded. The results of the regression analysis for the other variables are shown in Table 5.

One characteristic of the final concept map – proposition accuracy (the number of correct propositions) – was found to be a significant predictor ($p < .05$) of the post-test score.

Students were encouraged but not obliged to change their own concept maps after giving feedback. Overall, 47 % of students did so (53 % in the low prior knowledge group, 40 % in the average group, and 60 % in the high group). As the greatest learning gain occurred in the group of low prior knowledge students and as the final version of the concept map was graded for the analysis, it is interesting to see whether the

Table 4 Mean (SD) scores for characteristics of reviewers' learning products, by condition.

	Maximum score	Comment condition <i>n</i> = 46	Smiley condition <i>n</i> = 47	Total <i>n</i> = 93
Changing CM	1.00	.40 (.50)	.53 (.50)	.47 (.50)
CM proposition accuracy	n/a	3.22 (2.03)	3.26 (2.30)	3.24 (2.16)
CM salience	1.00	.60 (.31)	.56 (.32)	.58 (.31)
CM complexity	3.00	1.06 (0.50)	1.02 (0.53)	1.04 (0.51)
Quality of feedback given	5.00	2.46 (1.05)	2.80 (1.07)	2.63 (1.07)

Note: CM = concept map.

Table 5 Regression analysis of factors influencing post-test scores.

	<i>B</i>	<i>t</i>	<i>p</i>
Changing CM	−0.708	−1.565	.121
CM proposition accuracy	0.354	2.643	.010
CM complexity	−0.955	−1.676	.097
Quality of feedback given	0.261	1.215	.228

Note: CM = concept map.

Table 6 Percentage of students who changed their own concept maps in the low prior knowledge group, by condition.

	Comment condition (%) <i>n</i> = 10	Smiley condition (%) <i>n</i> = 5	Whole group (%) <i>n</i> = 15
Changed CM	70	20	53
Did not change CM	30	80	47

Note: CM = concept map.

intervention stimulated them to change their own product. Descriptive statistics for this group suggest that students in the comment condition changed their concept maps more often than those in the smiley condition (Table 6).

4. Conclusion and discussion

The feedback-giving component of the peer assessment process has been studied less than the feedback-receiving component, although several studies have demonstrated that a reviewer learns from the reviewing process (Cho & Cho, 2011; Li et al., 2010; Lu & Zhang, 2012; Lundstrom & Baker, 2009; Patchan & Schunn, 2015). The goal of the current study was to contribute to better understanding of the effectiveness of giving feedback; in particular, we aimed at investigating which way of giving feedback contributes more to a reviewer's learning – giving comments or grading with smileys. The context of the study was rather unique, which distinguished it from many of the studies conducted so far. First, the study took place in the inquiry learning environment created for secondary school students. As giving feedback can be seen a challenging task for students, not many studies have targeted secondary school. Checking whether the results of other studies apply with this context and age group can contribute to better understanding of the feedback-giving process. Second, participants had to give feedback on concept maps, which are rather small-scale learning products compared to, for example, pieces of writing (e.g., Xiao & Lucking, 2008; Wooley et al., 2008). Investigating whether a brief feedback-giving moment can lead to any learning for a reviewer has practical value, as giving feedback on smaller-scale learning products can be implemented in a classroom context more easily than giving feedback on larger-scale products. Finally, having the feedback giving take place online and anonymously also aimed to eliminate the influence of social factors such as personal relationships and peer pressure, which can be very important, especially for the target age group.

The feedback support for both conditions was structured to follow the recommendations by Hattie and Timperley (2007) for effective feedback: the criteria for evaluating the product presented the desired state of the product and the reviewing part pointed to the problems and suggested the direction for improvement. Such direction was more explicitly present in the feedback from the commenting condition, while in the grading condition, the combination of criteria with grades would show what to improve.

An ANOVA was conducted to find the effect of condition on the learning result. Independent variables included condition and prior knowledge level. Even though the feedback-giving moment was rather

brief, statistically significant effects were found. Besides an unsurprising main effect for prior knowledge, a main effect for condition was found, as well as an interaction effect for condition and prior knowledge level. The main effect for prior knowledge showed that a higher pre-test score led to a higher post-test score. Even though looking at the trend for different prior knowledge groups suggested that this main effect was mostly caused by the low prior knowledge students, the effect was statistically significant. It demonstrated that providing feedback in the form of comments led to higher post-test scores than providing feedback in the form of grades (smileys).

The interaction effect demonstrated that the same way of giving feedback contributed differently to the learning results depending on the reviewer's level of prior knowledge. The observed trend suggested that giving feedback in the form of comments might be most beneficial for students with lower prior knowledge.

Finding the commenting condition to be more beneficial for reviewers' knowledge gain was in line with some of the previous studies (Lu & Law, 2012; Wooley et al., 2008). In terms of implications, the results may indicate that asking students to give comments leads to better learning results for the reviewers overall than having them just grade the products of other students. This can be especially useful in a situation with smaller-scale learning products, as commenting should not be too time-consuming then for students.

Finding that the way feedback is given works differently for different prior knowledge groups was a new result. Previous research on the role of domain knowledge in giving feedback (e.g., Alqassab et al., 2018b; van Zundert, Könings et al., 2012) suggested that students should have enough knowledge to give feedback. In our study, even the low prior knowledge group could learn from giving feedback by commenting, which means that this particular task was manageable for all knowledge groups. As the biggest difference in the post-test results was observed for the low prior knowledge students, this group in particular should be given the opportunity to comment instead of to grade.

These results were obtained in a very special context – inquiry learning – and might be different in a different context. However, the fact that they are in line with some other studies showing that commenting is more beneficial for reviewers' learning (e.g., Xiao & Lucking, 2008; Wooley et al., 2008) indicates that this trend can be found in different learning contexts, which makes it more generalizable. Moreover, finding that commenting even on a small-scale learning product could contribute to the reviewer's learning makes it more applicable to everyday school practice. Another conclusion regarding the practical usage of peer feedback is that it can be used not only for evaluation purposes, but also as a learning tool, as it triggers more learning for reviewers, which is in line with previous research (see, e.g., van Popta et al., 2017).

To have a deeper understanding of what triggers learning when giving feedback, a regression analysis was conducted to see which factors – such as the quality of reviewers' own concept maps and of the feedback given – contributed to the post-test results. The number of correct links in students' own final concept maps was found to be a significant predictor of post-test scores, which means that concept maps reflected the knowledge that students had about the domain. This finding aligns with the understanding of the role that concept maps play in learning. Being able to choose relevant terms and connect them properly reveals deeper understanding of the topic (see, e.g., Novak & Cañas, 2006). Following this direction may encourage using concept mapping in secondary schools more often.

In the low prior knowledge group, where the most learning happened, the percentage of students who changed their concept maps after giving feedback was higher in the commenting condition than in the grading condition. Going back and changing one's own concept map might be a mechanism that triggers better understanding of the topic. Based on the results, it can be assumed that providing comments stimulates students to do so more than grading. These results are aligned with the findings by Harrison, Gerard, and Linn (2018), who observed

that providing a critique leads to re-working of students' own products more than just being instructed to re-work them. In other words, giving critical feedback to peers' pieces of work stimulates a more critical attitude toward one's own product, which, in turn, prompts revisiting it. Just being encouraged or even being provided with some hints about revisiting one's own products had a smaller effect. This may mean that students appreciate a more independent way of learning, in which they make the decision about improving their product themselves, rather than being directed to do so.

There are several limitations of the current study. First, even though the sample size was sufficient for the intervention with two conditions, the low and high prior knowledge groups were quite small and not very evenly distributed between the conditions. This limits the generalizability of the results for those groups. Second, the intervention was quite brief, which could make the effect less obvious. Moreover, participants were exposed to at least two new and potentially challenging tasks during the same lesson – concept mapping and giving feedback to peers. These two factors together – brief intervention and no prior experience in giving feedback on peers' concept maps – created a rather unique combination, which means that the results may not be same in a different context.

The observed trends suggest that in giving feedback, there is no “one size fits all” solution. Therefore, further exploring the phenomenon of giving feedback seems worthwhile. One direction for future research can be checking on the observed trends with a larger sample to see if the differences between groups with different levels of prior knowledge become more obvious. Conducting such a study with a larger sample in different countries would also include sociocultural aspects that were not the focus in the present study. Another direction can cover studying different aspects involved in giving feedback and their connection with the reviewer's prior knowledge level, which can lead to deeper understanding of the phenomenon and suggestions for practice. For example, the first step in giving feedback, namely, developing the assessment criteria, could be manipulated. The focus could be to investigate whether producing their own criteria or using ones that are given is more beneficial for reviewers' learning.

Compliance with ethical standards

The approval of the Ethical Committee of the Faculty of Behavioral, Management, and Social Sciences (BMS) University of Twente was obtained prior to the data collection.

Declaration of Competing Interest

The Authors declare that there is no conflict of interest.

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.

References

- 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>.
- Cao, Z., Yu, S., & Huang, J. (2019). A qualitative inquiry into undergraduates' learning from giving and receiving peer feedback in L2 writing: Insights from a case study. *Studies in Educational Evaluation*, 63, 102–112. <https://doi.org/10.1016/j.stueduc.2019.08.001>.
- 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>.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20, 328–338. <https://doi.org/10.1016/j.learninstruc.2009.08.006>.
- de Jong, T., Sotiriou, S., & Gillet, D. (2014). Innovations in STEM education: The Go-Lab federation of online labs. *Smart Learning Environments*, 1, 1–16. <https://doi.org/10.1186/s40561-014-0003-6>.
- Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Abingdon: Routledge Falmer.
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37, 16–55.
- Harrison, E. J., Gerard, L., & Linn, M. (2018). Encouraging revision of scientific ideas with critique in an online genetics unit. In J. Kay, & R. Luckin (Eds.), *Rethinking learning in the digital age: Making the learning sciences count 13th International Conference of the Learning Sciences (ICLS) 2018* (Volume 2, pp. 816–823). London, UK: International Society of the Learning Sciences.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <https://doi.org/10.3102/003465430298487>.
- Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics, Vol. 2, Reading, writing and language processing* (pp. 176–240). Cambridge, UK: Cambridge University Press.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41, 212–218. https://doi.org/10.1207/s15430421tip4104_2.
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education*, 45, 193–211. <https://doi.org/10.1080/02602938.2019.1620679>.
- Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41, 525–536. <https://doi.org/10.1111/j.1467-8535.2009.00968.x>.
- 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>.
- Lu, J., & Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A path model. *Journal of Educational Computing Research*, 46, 313–333. <https://doi.org/10.2190/EC.46.3.f>.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18, 30–43. <https://doi.org/10.1016/j.jslw.2008.06.002>.
- Novak, J. D., & Cañas, A. J. (2006). *The theory underlying concept maps and how to construct them (Technical report IHMC CmapTools 2006-1)*. Pensacola, FL: Florida Institute for Human and Machine Cognition.
- Patchan, M. M. (2011). *Peer review of writing: Learning from revision using peer feedback and reviewing peers' texts [Unpublished doctoral dissertation]*. PA: University of Pittsburgh, Pittsburgh.
- 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., ... Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>.
- Phillips, F. (2016). The power of giving feedback: Outcomes from implementing an online peer assessment system. *Issues in Accounting Education*, 31, 1–15. <https://doi.org/10.2308/iace-50754>.
- 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, 260–278. [https://doi.org/10.1002/1098-2736\(200102\)38:2<260::AID-TEA1005>3.0.CO;2-F](https://doi.org/10.1002/1098-2736(200102)38:2<260::AID-TEA1005>3.0.CO;2-F).
- Sluijsmans, D. M. A. (2002). Student involvement in assessment. *The training of peer assessment skills [Unpublished doctoral dissertation]*. The Netherlands: Open University of the Netherlands.
- Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, 20, 265–269. <https://doi.org/10.1016/j.learninstruc.2009.08.002>.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276. <https://doi.org/10.3102/00346543068003249>.
- 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 Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R.-J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24–34. <https://doi.org/10.1016/j.edurev.2016.10.003>.
- van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies*, 38, 541–557. <https://doi.org/10.1080/03055698.2012.654920>.
- van Zundert, M. J., Sluijsmans, D. M., Könings, K. D., & van Merriënboer, J. J. (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>.
- Wooley, R., Was, C., Schunn, C. D., & Dalton, D. (2008). The effects of feedback elaboration on the giver of feedback. *Paper Presented at the 30th Annual Meeting of the Cognitive Science Society*.
- Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11, 186–193. <https://doi.org/10.1016/j.iheduc.2008.06.005>.
- Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment and Evaluation in Higher Education*, 45, 372–386. <https://doi.org/10.1080/02602938.2019.1644603>.