Giving feedback to peers: How do students learn from it?

Natasha Dmoshinskaia

GIVING FEEDBACK TO PEERS: HOW DO STUDENTS LEARN FROM IT?

Natasha Dmoshinskaia

GIVING FEEDBACK TO PEERS: HOW DO STUDENTS LEARN FROM IT?

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Friday the 5th of March 2021 at 14:45 pm

by

Nataliia Glebovna Dmoshinskaia
born on the 8th of May 1978
in Leningrad, Russia

This dissertation has been approved by:

Supervisor: Prof. dr. A. J. M. de Jong

Co-supervisor: Dr. A.H. Gijlers

Cover design: Casper de Jong

Printed by: Ipskamp Printing

Lay-out: Sandra Schele

ISBN: 978-90-365-5117-5

DOI: 10.3990/1.9789036551175

Graduation committee

Chair:	Prof. dr. T.A.J. Toonen ▪ University of Twente
Supervisor:	Prof. dr. A.J.M. de Jong ▪ University of Twente
Co-supervisor:	Dr. A.H. Gijlers ▪ University of Twente
Committee members:	Prof. dr. S.E. McKenney ▪ University of Twente
	Prof. dr. P.C.J. Segers ▪ University of Twente
	Prof. dr. M. Pedaste ▪ University of Tartu
	Prof. dr. Z. Zacharia ▪ University of Cyprus
	Dr. P. Papadopoulos ▪ University of Twente

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 Program. This thesis 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.

ACKNOWLEDGEMENTS

Any road we take brings us to people – people who travel with us, people who inspire us, people who teach us, people who guide us, people who encourage us, people who make us believe in ourselves, people who laugh with us. These acknowledgements is my way to think about the four-year road of my PhD and to thank people for the role they played in this journey.

I would like to start by thanking my supervisors Ton de Jong and Hannie Gijlers. Ton, I really appreciate your believing in me and offering me an opportunity to do a PhD at this department. Thank you for giving me freedom to choose my studies and guidance to write about them; for giving me a chance to be a team member in such a big project as Next-Lab and to have my own responsibilities in it; for supporting me after article rejections and showing me the way to use them to improve further. Hannie, I am very grateful that you joined my ongoing project and helped me to structure it; that I could share my doubts with you and always get a supportive reply; that your feedback helped me to see things from a different angle and to carry on. Thanks to both of you, I have learned a lot during these four years, both professionally and personally.

My project would not have been the same without the Go-Lab environment, literally as well as metaphorically. Meeting and working together with such enthusiastic, devoted and experienced colleagues as the Next-Lab team was not only very educative but also very inspiring. Thank you for showing me a broader picture of research and practice outside my PhD. I would like to express a special thanks to Jakob for developing the tool and being open to all my requests, questions and troubleshooting. It made my journey very comfortable from a technical point of view.

I would like to thank all the teachers and students, both in the Netherlands and in Russia, who participated in my studies and made it all happen. Jeannet, Robin, Gerdi, Joanne, Erny, Olga, Natalia and many others I am not naming here, thank you for being open, enthusiastic, supportive and helpful; for thinking along and for taking a personal interest in what I am doing. It was incredibly important and rewarding for me to be in your classrooms, to see the process and to celebrate the results.

I felt extremely lucky to do my PhD journey at the IST department. I am very thankful to the colleagues for being supportive professionally and personally; for creating food for thought during research meetings and sharing real food during ‘uitjes’; for challenging my skills (including language ones) and being there to answer my questions. Thank you, Sandra, for all the help and talks, and ideas, and Christmas tree decorating moments. Thank you, Alieke, for supporting my UTQ trajectory and for sharing my ups and downs on the way to this moment. Thank you all for big and small things I will remember. I would also like to thank one person outside the department: Emily Fox, your critical, relevant and through comments always made my texts a better piece to read.

My fellow-PhDs, how many different roles you played: a psychological support group, helpful reviewers, companions in laugh, and so many other. Thank you very much for travelling along; for sharing; for making the road look manageable. Xiulin, it was great to be your roomie, your hard work encouraged me and your readiness to help meant a lot to me. My great paranymphs – Elise and Johannes – thank you for helping with more and less serious choices; for discussing my statistical or/and philosophical questions; and simply for being there for me.

My work would not have been completed if I had not felt loved, supported, appreciated, encouraged and believed in. My dear husband, Martin, your faith in me makes me stronger and your care allows me to act. Ik ben zo gelukkig met en zo dankbaar voor alle (schoon-)familie en vrienden die ik hier heb ontmoet voor hun steun, zorg en begrip. Jullie hebben me hier thuis laten voelen en het hielp enorm.

And the last but definitely not the least, I would like to thank my family and friends in Russia. Мои дорогие родители, спасибо, что поддерживаете меня, несмотря на то, что моё отсутствие в России для вас нелегко. Ваша любовь и забота преодолевает все границы и расстояния. Спасибо моим дяде, тёте и братьям за участие и живой интерес к тому, что я делаю. Спасибо моим друзьям (Яне, Ане, Илье, Валере, Юре и всем остальным) за такие разные и такие нужные мне разговоры, советы, наставления, шутки, а главное, за поддержку в тяжелые моменты и радость в хорошие. Без моей семьи и друзей я бы не оказалась там, где я сейчас, спасибо вам за это!

Natasha

TABLE OF CONTENTS

Chapter 5: Type of reviewed products

A study of the effect of the different types of the reviewed products

Chapter 6: General discussion

Chapter 7: English summary

Chapter 8: Nederlandse samenvatting

I

General introduction

INTRODUCTION

Every day we interact with other people. Such interaction may include giving and receiving feedback; moreover, people who give feedback to us may not be the same people we give feedback to. This feedback may take a whole spectrum of forms and formats from official job evaluations to giving likes in social media. And when we assess other people, we think we do good for them. But is this the only good that can originate from assessing our peers? When we give feedback, can we maybe also learn from this ourselves? And if we can, how can we include peer assessment into education so that it benefits students, especially the ones giving feedback, the most?

This dissertation explores the learning potential of peer assessment, in particular one part of this process: giving feedback to peers. Studying giving feedback can help to understand two things: its role and added value in education and the way to organize it so that it contributes to students' learning the most.

PEER ASSESSMENT

Peer assessment as a teaching method, is gaining popularity among teachers at different levels of education (e.g., L. Li & Grion, 2019; Tsivitanidou, Constantinou, Labudde, Rönnebeck, & Ropohl, 2018). At first, it was mainly seen as a replacement of or an addition to teacher's assessment, making peer assessment especially useful in a situation of a big group of students. Such approach inspired a large body of research investigating the validity and reliability of peer's assessment compared to an expert one (e.g., Hovardas, Tsivitanidou, & Zacharia, 2014; Patchan, Schunn, & Clark, 2018; Zhang, Schunn, Li, & Long, 2020).

Nowadays, however, peer assessment is more and more seen as an independent learning activity (e.g., van Popta, Kral, Camp, Martens, & Simons, 2017). So not surprisingly, now more research is being carried out investigating different factors that can influence learning originating from peer assessment. Both parts constituting peer assessment – giving feedback to peers and receiving feedback from them – have been shown to influence learning positively. For example, being involved in peer assessment helps students to acquire more responsibility for their own learning, develop evaluative skills and improve their performance (e.g., H. Li, Xiong, Hunter, Guo, & Tywoniw, 2020).

When investigating giving and receiving peer feedback separately, several studies demonstrated that giving feedback can be even more beneficial for students' learning than receiving feedback (e.g., Ion, Sánchez-Martí, & Agud-Morell, 2019; L. Li & Grion, 2019; Phillips, 2016). Thus, studying giving feedback would mean investigating students' learning and the ways to facilitate such learning even more than it is done now. Therefore, the focus of this thesis is to go deeper into the peer feedback-giving process to understand how to organize it efficiently, i.e., to maximize learning originating from giving feedback.

GIVING FEEDBACK TO PEERS

As it was mentioned above, giving feedback to peers can lead to learning of the feedback provider. In our studies, we name feedback providers ‘reviewers’. The choice of the term reviewer over other terms, for example ‘assessor’, was made to emphasize the formative purpose of giving feedback in our studies. This purpose encourages students to think of the way of improving a reviewed product rather than only spotting mistakes, which leads to more reflection about the topic compared to a summative assessment task.

Several researchers looked into the process of giving peer feedback in more detail. Sluijsmans (2002) suggested the following model for this process (Figure 1.1).

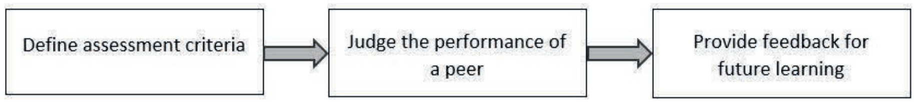

Figure 1.1. A three-step model of giving feedback to peers (by Sluijsmans, 2002)[1]

Learning from giving feedback can be attributed to the fact that reviewers are cognitively involved with the topic as they are reviewing peers’ products or performance. This, in turn, triggers thinking about the important characteristics of the product or performance and, as a result, learning. Following the model, reviewers are going through several steps, each of which includes sub-steps and sub-skills (Sluijsmans, 2002). First, they need to understand given assessment criteria or come up with their own ones. This step is supposed to lead to better understating of the key characteristics of the to-be-reviewed product. Second, they need to compare a reviewed product against assessment criteria, which should to bring more topic understanding as well as develop their evaluative skills. Finally, reviewers need to suggest ways to correct mistakes or improve the quality of the product. This step should encourage reviewers to apply their (gained) understanding of the topic to provide a recommendation, which together with previous steps leads to learning.

Looking for the ways to increase learning from giving feedback to peers, one can study the factors that influence each step of the feedback-giving process. Finding the optimal combination of these factors can provide the basis for practical recommendations. Such factors can include different assessment criteria, different

[1] In the work by Sluijsmans the term ‘assessment criteria’ was used to introduce criteria to compare peers’ products against and give feedback based on this comparison. In all our studies, we use the term ‘assessment criteria’ for criteria to give feedback on peers’ products.

types of reviewed product or different way to give feedback. Being given assessment criteria or developing own criteria may lead to focusing on different product's characteristics while giving feedback, which, in turn may lead to different learning. The type of the product to give feedback on could influence the process of giving feedback as different products' types may stimulate different interaction with the learning material and, thus, different learning outcomes. Finally giving feedback by commenting or by grading peers' work might require different cognitive involvement and result in different learning. Investigating these factors was the goal of this dissertation, with more details given later in the dissertation outline.

There was another factor to consider. Several studies showed that students' prior knowledge can influence the quality of given feedback, for example, the number of spotted errors, or the amount of provided recommendations (e.g., Alqassab, Strijbos, & Ufer, 2018; Patchan, Hawk, Stevens, & Schunn, 2013; van Zundert, Könings, Sluijsmans, & van Merriënboer, 2012). Assuming that quality of feedback may influence students' learning, reviewers' prior knowledge was taken into account in each conducted study.

In our studies, participants gave feedback on concept maps. Concept maps were chosen as they allow students to think about key concepts of the topic and relationships between them, which contributes to deeper understanding (e.g., Novak, 2010). Moreover, students who reviewed concept maps were shown to learn more than students who did not participate in reviewing (e.g., Chen & Allen, 2017). Therefore, using concept maps as a to-be-reviewed product might lead to more learning originating from reviewing it than when reviewing other products, which can make the influence of other factors more visible.

INQUIRY LEARNING AND GO-LAB ECOSYSTEM

All studies were conducted in an online inquiry learning environment, which distinguish the current research from most done in the field of giving feedback to peers. Using such environments created a unique context for the studies.

Inquiry learning supports students' exploration of the topic in a way that resembles a scientific investigation. Students take a role of scientists and answer a research question by checking hypotheses through an experiment. Such a process usually includes several steps and is referred to as an inquiry learning cycle. Pedaste et al. (2015) analyzed the existing models of inquiry learning cycle and argued that most of them include the same steps but call them different names. The authors summarized steps from different models with the same content and suggested a cycle that included five steps. According to Pedaste et al., inquiry learning includes the following stages: orientation, conceptualization, investigation, conclusion and discussion. This is the model that was used in the dissertation.

Following the inquiry learning cycle, students explore a topic by creating hypotheses, testing them in an online lab and drawing conclusions based on the experiments. The last phase of the cycle (discussion) is particularly suitable for a peer feedback activity as it helps students to reflect on peers' and their own learning products that have been created during the inquiry.

To be beneficial for students' learning inquiry learning should be guided (e.g., de Jong, 2006; de Jong, et al., in press; Lazonder & Harmsen, 2016; van Riesen, 2018). The Go-Lab ecosystem (see www.golabz.eu) aims at supporting online inquiry learning with the help of Inquiry Learning Spaces (ILSs). Such spaces facilitate students' following an inquiry cycle and they can be created using the ecosystem repository of online labs, supporting tools and specifically developed scenarios integrating relevant instruments in each phase of inquiry. Lesson materials (ILSs) for all studies were designed using the Go-Lab ecosystem. Feedback to peers was given with a help of a special Peer Assessment tool that allowed students to see reviewed products, assessment criteria and give feedback anonymously.

A combination of a feedback-giving activity and an inquiry learning lesson had two-folded benefits. On the one hand, giving feedback in an inquiry context can be applied more naturally than with a traditional instruction method because giving feedback to peers is a part of a scientific research cycle. In a real-life situation, scientists give each other feedback at numerous occasions, like participating at a round table or peer review an article for a scientific journal. Critiquing peers' products and providing feedback on them helps students in developing scientific reasoning and conceptual understanding (e.g., Dunbar, 2000; Friesen & Scott, 2013). On the other hand, giving feedback to peers can add value to the instructional method of inquiry learning. According to the categorization suggested by Chi (2009) an inquiry learning lesson is a constructive activity as students explore the topic themselves. Including a full peer-assessment cycle would turn the inquiry process into an interactive activity, as students would have a chance to discuss and change their products in real time. Having only a feedback-giving activity is an intermediate step that can be defined as semi-interactive as students get involved with the task more than without giving feedback. Therefore, peer feedback and an inquiry learning lesson could mutually enrich each other and lead to more learning compared to the situation when only one method is used.

Problem statement and dissertation outline

As mentioned at the beginning of this chapter, peer assessment is viewed as a learning activity with giving feedback potentially contributing to learning more than receiving feedback. As giving feedback itself is a process consisting of several steps, different factors can influence each of these steps and, as a result, learning originating from it. Thus, studying these factors can contribute to understanding of the process of giving feedback better and to organizing it in the way that benefits learning the most. Moreover, taking into account reviewers' prior knowledge can help to understand if different prior knowledge may lead to differences in given feedback and if differences in given feedback may lead to differences in learning.

Therefore, the overall aim of the dissertation was to understand reviewers' learning and discover how to increase it by designing the feedback-giving process in a particular way. To do so several studies were conducted, with each study focusing on one particular factor that could influence the process of giving feedback. Using the model of the peer feedback-giving process suggested by Sluijsmans (2002), the following studies were carried out (Figure 1.2).

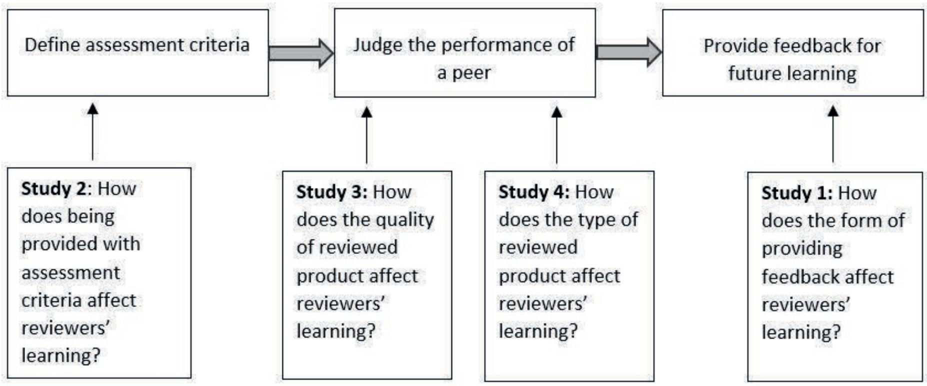

Figure 1.2. Overview of the studies

In their paper about the concept of peer assessment, Strijbos and Sluijsmans (2010) identified several gaps in the research in this field. In particular, they suggested having more (quasi-) experimental studies focusing on a formative rather than a summative component of peer assessment and its effect on learning. A more recent meta-analysis of the trends in the studies of peer assessment conducted by Fu, Lin, and Hwang (2019) demonstrated that this situation has not changed dramatically. The authors emphasized the necessity to conduct research using peer assessment with learning rather than assessing purposes, for example, for developing students' conceptual understanding by using peer assessment. They also promoted research on the implementation of peer assessment in elementary and secondary schools and not only for higher education. The present dissertation aimed at filling in (to some extent) the gaps indicated by Fu et al., and Strijbos and Sluijsmans. First, the focus of the studies was on the influence of different factors that might increase learning of a feedback provider. As students gave feedback on concept maps created during an inquiry learning lesson, giving feedback was expected to contribute to conceptual learning in the first place. Second, all conducted studies were experimental. And finally, the target group for the experiments was secondary school students.

The dissertation has the following outline:

Study 1 aimed at comparing two ways of giving feedback on peers' concept maps using given assessment criteria: write comments and grade with smileys. Learning results and products of peer reviewers were compared. This study is presented in Chapter 2.

Study 2 aimed at investigating the role that being provided with assessment criteria plays in the feedback-giving process. For that, two conditions were compared: one giving feedback on peers' concept maps using given assessment criteria and the other had to come up with their own assessment criteria. Learning results and products of peer reviewers were compared. This study is presented in Chapter 3.

Study 3 aimed at examining the effect of the level of the reviewed product on reviewers' learning. For that, three conditions were compared giving feedback on peers' concept maps of various quality: the first group reviewing concept maps of a lower quality, the second groups reviewing concept maps of a mixed quality, and the third group reviewing concept maps of a higher quality. Learning results and products of peer reviewers were compared. This study is presented in Chapter 4.

Study 4 aimed at exploring if different type of reviewed products could stimulate different learning of peer reviewers. Giving feedback on two different peers' products was compared: concept maps and test answers. Learning results and products of peer reviewers were compared. This study is presented in Chapter 5.

General *discussion* summarizing the finding of all studies and providing conclusions of the dissertation is presented in Chapter 6.

References

Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*, 11-30. doi:10.1007/s10212-017-0342-0

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

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73-105. doi:10.1111/j.1756-8765.2008.01005.x

de Jong, T. (2006). Scaffolds for scientific discovery learning. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 107-128). Boston: Elsevier.

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., . . . Zacharia, Z. C. (in press). Understanding teacher design practices for digital inquiry-based science learning: The case of Go-Lab. *Educational Technology Research & Development*.

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology, 21*, 49-58. doi:10.1016/S0193-3973(99)00050-7

Friesen, S., & Scott, D. (2013). *Inquiry-based learning: A review of the research literature*. Paper prepared for the Alberta Ministry of Education.

Fu, Q.-K., Lin, C.-J., & Hwang, G.-J. (2019). Research trends and applications of technology-supported peer assessment: A review of selected journal publications from 2007 to 2016. *Journal of Computers in Education, 6*, 191-213. doi:10.1007/s40692-019-00131-x

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education, 71*, 133-152. doi:10.1016/j.compedu.2013.09.019

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*, 124-138. doi:10.1080/02602938.2018.1484881

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*, 681-718. doi:10.3102/0034654315627366

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*, 193-211. doi:10.1080/02602938.2019.1620679

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

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

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41*, 381-405. doi:10.1007/s11251-012-9236-3

Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: Examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education, 43*, 2263-2278. doi:10.1080/03075079.2017.1320374

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. doi: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. doi:10.2308/iace-50754

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.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269. doi:10.1016/j.learninstruc.2009.08.002

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. *European Journal of Psychology of Education, 33*, 51-73. doi:10.1007/s10212-017-0341-1 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. doi:10.1016/j.edurev.2016.10.003

van Riesen, S. A. N. (2018). *Inquiring the effect of the experiment design tool: Whose boat does it float?* (unpublished doctorate dissertation). University of Twente, The Netherlands.

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. doi:10.1080/03055698.2012.654920

Zhang, F., Schunn, C., Li, W., & Long, M. (2020). Changes in the reliability and validity of peer assessment across the college years. *Assessment & Evaluation in Higher Education*, *45*, 1073-1187. doi:10.1080/02602938.2020.1724260

2

Feedback method

A study of the effect of the method of giving feedback

This chapter is based on:

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (in press). Learning from reviewing peers' concept maps in an inquiry context: Commenting or grading, which is better? *Studies in Educational Evaluation*.

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.

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 et al., 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 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.

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 (2018a), 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 (2018b), 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 Merrriënboer (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.

Research questions

Several studies (Lu & Law, 2012; Wooley et al., 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?

METHOD

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).

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 2.1. 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.

Table 2.1
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?

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.

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 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 Figure 2.1). 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 Figure 2.2.
- 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.

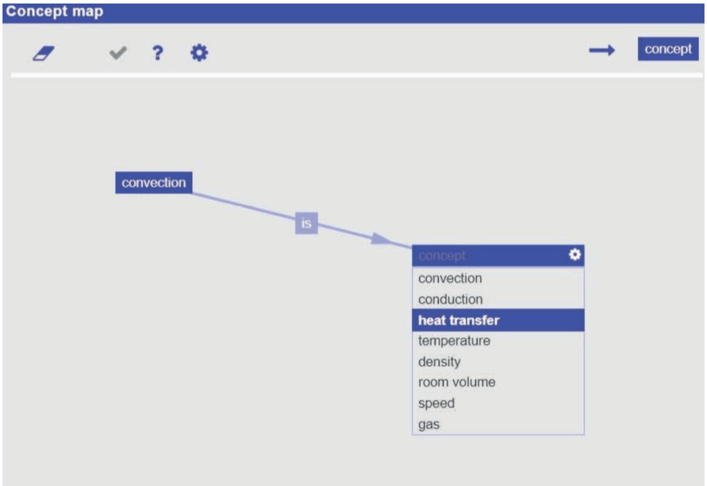

Figure 2.1. View of the Concept Mapper tool (translated)

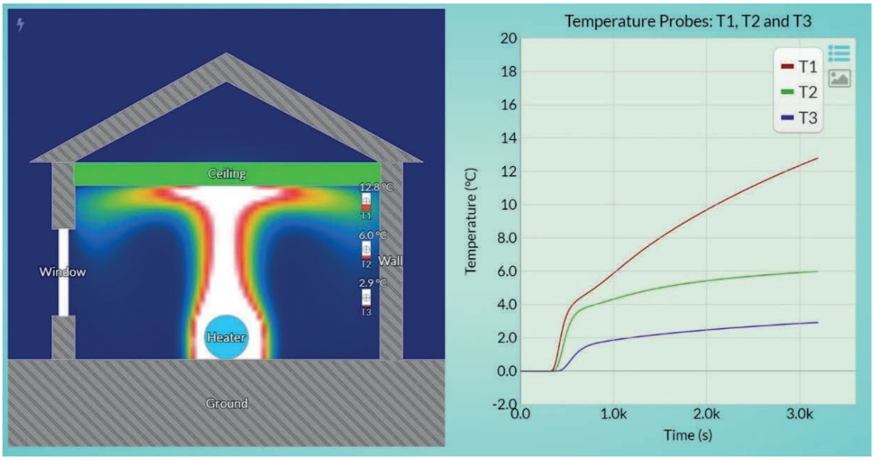

Figure 2.2. View of the online lab "Vertical temperature gradients".

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 Figure 2.3.

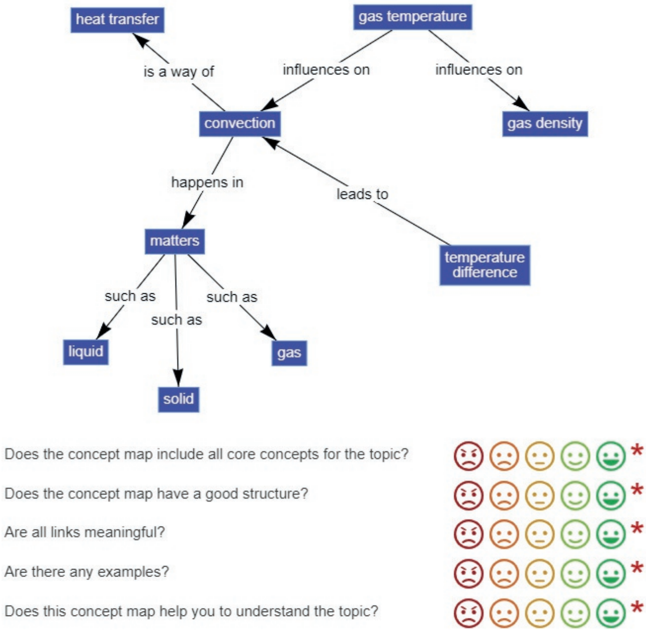

Figure 2.3. View of the peer assessment tool with a higher quality concept map, for the smiley condition (translated)

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 2.2 for an example).

Table 2.2
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

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 minutes) 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 minutes) 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 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.

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.

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.

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 2.3.

Table 2.3
Test scores (max score of 10) by prior knowledge level and condition

	Low prior knowledge		**Average prior knowledge**		**High prior knowledge**		**Total**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Comment condition	*n* = 10		*n* = 30		*n* = 6		*n* = 46	
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	*n* = 5		*n* = 28		*n* = 14		*n* = 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	*n* = 15		*n* = 58		*n* = 20		*n* = 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

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, paired *t*-tests for pre- and post-test scores were used to check if students learned during the experiment. The results showed that students did learn: post-test scores were significantly higher than pre-test scores, both for the whole group [$t(92) = 3.43$, $p < 0.01$, Cohen's $d = 0.36$] and in each condition [$t_{COMMENT}(45) = 2.79$, $p < 0.01$, Cohen's $d = 0.41$; $t_{SMILEY}(46) = 2.02$, $p < 0.05$, Cohen's $d = 0.29$].

Fourth, 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 [$F(2,87) = 19.32$, $p < .001$, $\eta_p^2 = .31$] and condition [$F(1, 87) = 5.84$, $p = .018$, $\eta_p^2 = .06$]. 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 effect, a significant interaction effect between condition and prior knowledge level was also found, $F(2, 87) = 4.19$, $p = .018$, $\eta_p^2 = .09$ (Figure 2.4). This means that conditions worked differently for students with different levels of prior knowledge.

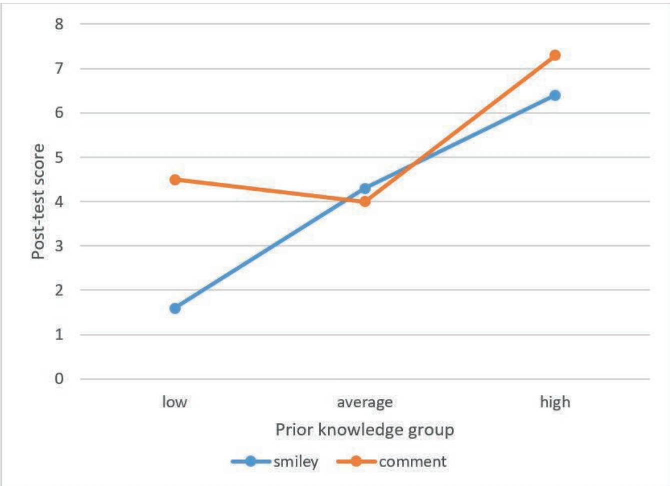

Figure 2.4. Post-test scores per prior knowledge group

To further specify this interaction effect, a separate analysis was conducted for each prior knowledge group. As the low and high ability 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 [$H_{\text{LOW}}(1) = 4.91$, $p = .027$, $H_{\text{AVERAGE}}(1) = 0.22$, $p = .639$, $H_{\text{HIGH}}(1) = 1.26$, $p = .261$].

Fifth, to obtain better understanding of the relation between the students' learning process and their learning performance, a regression analysis was performed, with the concept map quality scores and the quality of 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 2.4. The complete overview of the coding procedure was given in the analysis section.

Table 2.4
Scores for characteristics of reviewers' learning products, by condition

	Maximum score	**Comment condition**		**Smiley condition**		**Total**	
		M	*SD*	*M*	*SD*	*M*	*SD*
		n = 46		*n* = 47		*n* = 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 2.5
Regression analysis of factors influencing post-test scores

	B	***t***	***p***
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.

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 2.5.

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 intervention stimulated them to change their own product. Descriptive statistics suggest that students in the comment condition changed their concept maps more often than those in the smiley condition (Table 2.6).

Table 2.6
Percentage of students who changed their own concept maps in the low prior knowledge group, by condition

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

Note. CM = concept map.

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).

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.

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. doi: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. doi: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. doi:10.1016/j.stueduc.2019.08.001

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338. doi:10.1016/j.learninstruc.2009.08.006

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*, 629-643. doi:10.1007/s11251-010-9146-1

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. doi: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, 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. doi: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. doi: 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 & Evaluation in Higher Education, 45*, 193-211. doi: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. doi: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*, pp. 257-275. doi: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. doi: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. doi: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]. University of Pittsburgh, Pittsburgh, PA.

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. doi: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. doi: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. doi: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. doi: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]. Open University of the Netherlands, The Netherlands.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269. doi:10.1016/j.learninstruc.2009.08.002

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, pp. 249-276. doi: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. doi: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. doi: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. doi: 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. doi:10.1080/01443410.2011.626122

Wooley, R., Was, C., Schunn, C. D., & Dalton, D. (2008, July). *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. doi: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 & Evaluation in Higher Education, 45*, 372-386. doi:10.1080/02602938.2019.1644603

3

Providing assessment criteria

A study of the effect of providing assessment criteria

This chapter is based on:

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (in press). Giving feedback on peers' concept maps in an inquiry learning context: The effect of providing assessment criteria. *Journal of Science Education and Technology*. doi:10.1007/s10956-020-09884-y

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.

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, Xiong, Hunter, Guo, & Tywoniw, 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 & MacArthur, 2011; Ion, Sánchez Martí, & Agud Morell, 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. The results of the study by Canty, Seery, Hartell, and Doyle (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, Zacharia, and Hovardas (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 & De Wever, 2015; Sluijsmans, Brand-Gruwel, & van Merriënboer, 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, Romero, and Strijbos (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, Körndle, and Narciss (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, van Zundert, and van Merriënboer (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, Merry, and Reiling (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 & 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, Levy, & Century, 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, Xiong, Hunter, Guo, and Tywoniw (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, Coutinho, & Flores, 2016).

Research questions

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 & 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 & 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 3.1. Even though the assessment criteria were not domain-specific, the reviewed concept maps were, which led to giving domain-specific feedback.

Table 3.1
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 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 Figure 3.1.

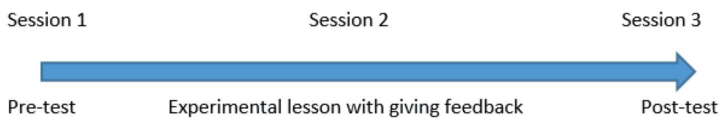

Figure 3.1. Timeline of the experiment

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 though 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 Figure 3.2) 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 Figure 3.3) 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.

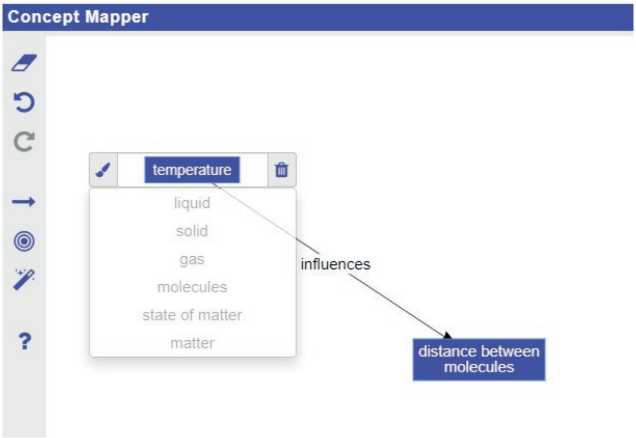

Figure 3.2. View of the Concept Mapper tool (translated from Russian)

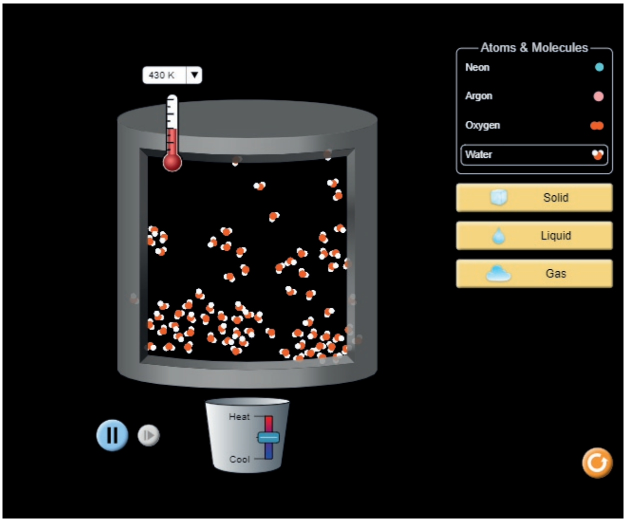

Figure 3.3. View of the online lab (English version, same as the Russian version used in the lesson). Images by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY 4.0 https://phet.colorado.edu

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 Figures 3.4 and 3.5, 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 & 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.

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 H_2O be at the temperature of 0° C at normal pressure?"; and "Do molecules change during a phase transition? Why?".

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 Figures 4 and 5.

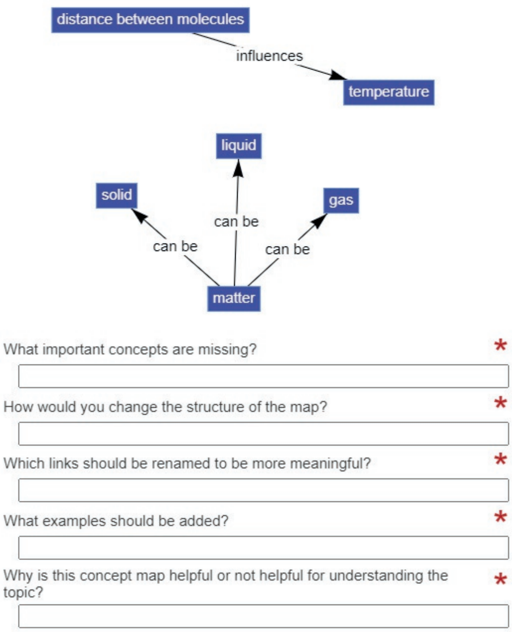

Figure 3.4. View of the peer assessment tool for the criteria condition, with the first concept map (translated from Russian)

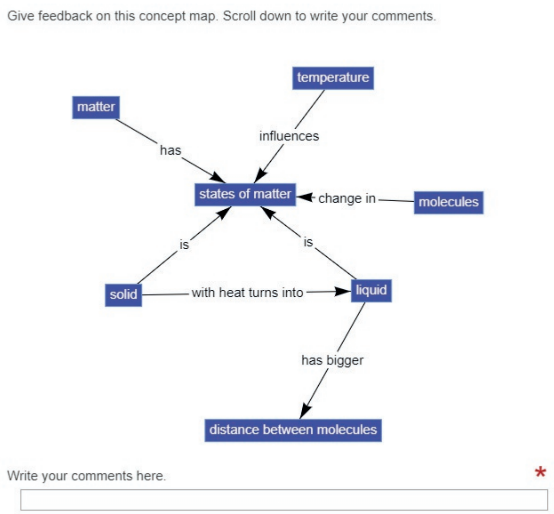

Figure 3.5. View of the peer assessment tool for the free condition, with the second concept map (translated from Russian)

Procedure

The experiment took place during regular school hours. The experimental lesson with the ILS was developed for one school lesson (45 minutes); 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.

Pre- and post-tests (about 15 minutes 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 .90 for pre-tests and .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,
- salience 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, Schultz, Li, and Shavelson (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 .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 .90, which meant almost perfect agreement.

RESULTS

The difference in pre-test scores between the conditions was not significant: $M_{FREE} = 3.41$ ($SD = 2.22$) and $M_{CRITERIA} = 3.31$ ($SD = 2.46$), $t(91) = 0.21$, $p = .83$ (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.

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 [$t(92) = 8.63$, $p < .01$, Cohen's $d = 0.89$] and in each condition [$t_{FREE}(43) = 6.74$, $p < .01$, Cohen's $d = 1.02$; $t_{CRITERIA}(48) = 5.61$, $p < .01$, Cohen's $d = 0.80$].

Table 3.2
Test scores by knowledge level and condition (maximum score is 12)

	Low prior knowledge		**Average prior knowledge**		**High prior knowledge**		**Total**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Free condition	*n* = 9		*n* = 27		*n* = 8		*n* = 44	
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	
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	
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

Descriptive statistics for all prior knowledge groups and both conditions are presented in Table 3.2.

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 [$F(2,87) = 8.24$, $p = .001$, $\eta_p^2 = .16$], 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 3.3. The coding schemes and maximum scores for these characteristics were discussed in the analysis section.

Table 3.3
Scores for feedback and concept map (CM) quality by condition

	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	.70	.30	.61	.37
Complexity of the CM	1.55	0.50	1.41	0.61
Changed CM after giving feedback	.57	.50	.41	.50

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 = .047), while for the criteria condition all variables were included (p = .020).

For both conditions, only the feedback quality was found to be a significant predictor of post-test scores [F_{FREE} (2, 39) = 3.30, $p < .05$, F_{CRITERIA} (5,43) = 3.04, $p < .05$]. 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 & 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 & 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, 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, Panadero, Schellens, & Raes, 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, Zacharia, and Hovardas (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: 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.

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. 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. doi: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. doi: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. doi.org/10.1016/S0193-3973(99)00050-7

Falchikov, N. (2004). Involving students in assessment. *Psychology Learning & Teaching, 3*(2), 102-108. doi: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. doi: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. doi: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. doi:10.1080/02602938.2018.1484881

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, *39*(10), 1774-1787. doi: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. doi: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. doi: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. 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. doi:10.1080/02602930050025006

Panadero, E., Romero, M., & Strijbos, 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. doi: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. doi: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. doi: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. doi: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. doi: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. 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]. 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. doi: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. doi: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, doi:10.1080/10494820.2013.834828

4

Quality of reviewed products

A study of the effect of the different quality levels of the reviewed products

This chapter is based on:

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (submitted). *Giving feedback on peers' concept maps as a learning experience: Does the quality of the reviewed concept maps matter?*

Abstract

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

INTRODUCTION

Peer assessment has proven to be not only a form of assessment, but also a powerful learning tool. Giving and receiving feedback from peers makes students reflect on important elements and characteristics of the products, thus leading to deeper thinking about the topic. Students who review peers' work and get feedback on their own create better learning products (see, e.g., Brill, 2016; Mulder, Baik, Naylor, & Pearce, 2014). For these reasons, peer assessment is gaining popularity, especially in online learning environments. A systematic literature review by Tenório, Bittencourt, Isotani, and Silva (2016) which focused on the phenomenon of peer assessment in online learning showed its effectiveness at various levels of education. The studies included in the review reported a positive effect of peer assessment on development of students' skills or knowledge. Moreover, according to a recent meta-analysis, a computer-mediated peer assessment leads to more learning than a paper-based one (Li, Xiong, Hunter, Guo, & Tywoniw, 2020). This is not surprising, as an online learning environment can facilitate in organizing a feedback-giving process more than a real classroom. For example, issues such as matching strategies, anonymity of a feedback-giving process or time limits can be tackled more easily in an online environment than in the real classroom. This means that students can be matched using different strategies even within one class, and they can give feedback anonymously and at their own pace. The combination of peer assessment and online learning environments can have advantages not only for implementing peer assessment but also for working with online environments. Students usually work individually in such environments, and a peer feedback activity can make the learning experience richer as students can see potentially different approaches to deal with the task when they give feedback to peers or receive feedback from them.

Providing feedback to peers is normally part of a peer-assessment process that consists of giving feedback to and receiving feedback from peers. Giving feedback is an active process and, thus, it requires students' active cognitive involvement. Such involvement may lead to a feeling of being more responsible for learning and more committed to it. For example, Ion, Sánchez-Martí, and Agud-Morell (2019) found that students perceived giving feedback as a learning experience more than receiving feedback, and that experience helped them to improve their performance. These results are not surprising, as giving feedback includes analyzing the product under review and applying one's existing knowledge to identify flaws and to suggest solutions. Moreover, giving feedback on peers' work can lead to reflection on one's own work and, thus, to more learning. However, the learning effects of providing feedback are potentially mediated by a few factors.

The first factor is the quality of the product that is being reviewed. Here results have indicated that students gave better feedback and learned more from reviewing a higher quality product than a lower quality product. In a study by Tsivitanidou, Constantinou, Labudde, Rönnebeck, and Ropohl (2018), secondary school students first had to construct a model of a physical phenomenon (color mixing), then give feedback on the models made by their peers, and finally work on their own model again, incorporating the feedback they received or just improving it based on their new ideas. Students who gave feedback on higher quality products changed their own products to align with the learning goals more than students who gave feedback on lower quality products. This means that feedback providers (reviewers) can learn by reviewing better examples produced by their peers. This is in line with the findings of Alqassab, Strijbos, and Ufer (2018a), who demonstrated that the quality of the reviewed product influences the quality of the feedback provided. In their study, preservice math teachers had to give feedback on peer solutions to a geometry proof. They were shown either an almost correct solution or one with many mistakes. The results showed that students who reviewed an almost correct solution gave more accurate feedback than those who reviewed a solution with mistakes. Moreover, students evaluating an almost correct solution had a better understanding of the geometric proof described in the solution. These results can be explained by the fact that to evaluate learning products, students must be competent enough to decide what is right and what is wrong. In this context, a low-quality learning product would require different comments and suggestions than a high-quality learning product. In other words, products with different levels of quality may require different levels of domain expertise from a reviewer. It can be easier to give feedback on a higher quality product, as reviewers may not be able to identify the problems in an incorrect product or suggest the right way to improve it.

This brings us to the second factor that plays a role in the reviewers' learning process: their prior knowledge or abilities. Several studies have investigated this issue. Cho and Cho (2011) looked at the process of giving peer feedback on technical research drafts. They analyzed the feedback given by university students, as well as the quality of the original and improved drafts. They found that reviewers' writing abilities and the level of the reviewed written pieces influenced the content of feedback. The first finding was that lower quality products received more critique mentioning weak points, whereas the higher quality products got mostly comments about their strong points. As for the prior abilities of the reviewers, reviewers with higher writing ability could identify more problems, but they gave the same amount of praise as reviewers with lower writing abilities. This means that initial abilities may matter in the process of giving feedback; the higher they are, the more problems can be identified, whereas strong points are equally easy for reviewers of different ability levels to spot. These findings are in line with the results of the study by Patchan and Schunn (2015), who further examined the relationship between the content of the feedback given and the reviewer's writing ability. The study was conducted with university students taking an introductory physiology course who were assigned to review peers' papers. This study showed that low-ability students were more likely to praise, while high-ability students were more likely to criticize by mentioning problems, but also suggested solutions. Moreover, there was an interaction effect found between the reviewers' ability and the quality of the reviewed product: high-ability reviewers could distinguish the product's quality by detecting more problems in lower quality products than in higher quality products, while low-ability reviewers provided the same amount of critique to both levels of product. Being able to provide constructive critique and solutions for the problems identified is connected with the learning process and can be seen in the reviewers' own learning performance. Yalch, Vitale, and Ford (2019) conducted a study with physiology students giving feedback on peers' written work. Their findings showed that the more critical feedback was given by students, the higher grades they received for their own written products.

Similarly to reviewers' writing abilities, it can be suggested that their prior knowledge may influence learning by giving feedback, especially in case of tasks requiring knowledge rather than skills (e.g., in science education). In terms of giving feedback, prior knowledge can influence the content of provided feedback, for example, reviewers with low prior knowledge may not be able to understand the product or come up with useful solutions (Alqassab, 2017). Such difference in reviewers' performance may lead to the difference in their learning. In the study by Patchan, Hawk, Stevens, and Schunn (2013), university students who had to review each other's work were paired differently according to their domain knowledge levels: either both students were at the same knowledge level or the two students had different knowledge levels. The findings of this study showed that low-knowledge and high-knowledge reviewers gave similar feedback on a good product (created by a high-knowledge student). However, the feedback on a weak product (created by a low-knowledge student) was different, with more identified problems coming from high-knowledge reviewers. This suggests that students can learn differently from reviewing products of different quality. Moreover, it is possible to assume that while higher-knowledge students can learn from different levels of products, lower-knowledge students might not learn much from a low-quality product, as they would not be able to identify problems and/or suggest solutions.

The type of learning product to give feedback on depends on the educational context. The current study was conducted in an online inquiry learning environment. Inquiry learning encourages students to explore a scientific topic and draw their own conclusions about the phenomenon being studied based on the results of experiments they have performed. Inquiry learning imitates the scientific research cycle and leads students through a (guided) exploration process. Inquiry learning includes several phases: orientation, hypothesis generation, experimentation, drawing a conclusion, and evaluation (Pedaste et al., 2015). During these phases, students can create several types of learning product, for example, hypotheses, concept maps, designs for experiments, conclusions, and so forth. Such products are usually small, as they are created in one or two lessons. The final phase – evaluation or discussion – is used to reflect on the process and the learning products that have been created, which makes this phase important for deeper learning. The activity of peer feedback, when positioned in this phase, can play a role in this reflective process, as it gives students an opportunity to evaluate peers' and their own products. Therefore, including giving feedback to peers in an inquiry lesson can contribute to better learning.

Research questions

The present study focuses on the role that the quality of reviewed products and reviewers' prior knowledge play in reviewers' learning. As the cited studies showed, the quality of the product to be reviewed may influence the quality and the content of the feedback provided. Moreover, as the cognitive processes for giving different types of feedback may differ, the quality of the product to be reviewed may influence the learning of a feedback provider. The current study investigated this issue in a secondary school context. Most of the studies mentioned looked at feedback given by university students on relatively large written products – essays and papers. It is not clear as yet if the trends are the same for a smaller type of learning product, such as concept maps, and a younger target population, such as secondary school children. Secondary school children were chosen as a target group for the following reasons. On the one hand, giving feedback to peers was found to be beneficial for learning of secondary school reviewers (e.g., Lu & Law, 2012; Wu & Schunn, 2020), and students of this age group were shown to be able to give meaningful feedback (Tsivitanidou, Zacharia, & Hovardas, 2011). On the other hand, there are only a few studies exploring giving feedback to peers at the secondary school level. Another difference from most of the above-mentioned studies is that they examined either the quality of the feedback provided or the reviewers' own learning products, while the current study took a broader look at the reviewers' learning. In our study, learning was analyzed through several parameters: post-test scores, the quality of students' own concept maps, and the quality of feedback given.

Altogether, this led to the following ***research questions***: Does the quality of reviewed products affect reviewers' learning when giving feedback to peers in a secondary school inquiry learning context? Is there a differential effect for students of different prior knowledge levels?

METHOD

Participants

The sample originally included 95 students from four classes at the same educational level (higher general secondary education, HAVO in Dutch) from one secondary school in the Netherlands, with average age 15.21 years old (*SD* = 0.48). Seventy-eight students (36 boys and 42 girls) participated in all three sessions and gave feedback to peers, so only their test results, concept maps and feedback given were used for the analysis. During the preliminary analysis an outlier was found (see the *Results* section for details), and this student was also excluded from the analysis, which gave a final sample of 77 students.

Within each class, students were assigned to one of three conditions:

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

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

Study design

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

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

The assessment criteria used in the study were not domain- or subject-dependent. Instead, they were specific to a particular type of a product (in our case, a concept map) and described its important characteristics. The criteria used were based on a study by van Dijk and Lazonder (2013) and are presented in the Table 4.1.

Table 4.1
Assessment criteria questions for giving feedback

1.	What important concepts are missing?
2.	Which links should be renamed to be more meaningful?
3.	What examples should be added?
4.	Why does or doesn't this concept map help with understanding the topic?

For ethical reasons (to avoid any dependency), test results or created learning products were not shared with the school teachers at the individual (student) level.

Materials

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

Students worked with an online inquiry learning environment that was created using the Go-Lab ecosystem (see www.golabz.eu). The lesson was constructed as an inquiry learning space (ILS) that facilitated students' going through an inquiry learning cycle, with experimenting in a virtual lab as the pivotal activity (de Jong et al, in press,). The ILS used in the current study followed a basic scenario for inquiry learning (see Pedaste et al., 2015) and consisted of the following phases:

- Orientation – in this phase the topic was introduced, as well as the idea of scientific research with experimenting in an online lab. The research question about the mechanism of phase transitions was set, and students' prior knowledge of the topic was refreshed with a short quiz covering the names of some phase changes and examples of them from a daily life.
- Conceptualization – in this phase the topic was narrowed to questions about the process of phase changing. Students could think about the most important concepts related to the topic by creating a concept map in a specially designed tool, the Concept Mapper (see Figure 4.1). The tool included some pre-defined concepts for the topic and names of links, organized in a drop-down menu. It also gave students the possibility of typing their own concepts and names of links. This combination provided some guidance, but still allowed freedom of learning.

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

- Investigation – in this phase the virtual lab (see Figure 4.2) was used for conducting an experiment and checking hypotheses. A brief introduction of the lab's settings was provided but students had the freedom to design the experiment themselves. In the lab, they could change the temperature of the matter to see what happened with its molecules. They could also draw correspondences between different stages of the temperature-changing process and the states of matter – solid, liquid, gas. Observations could be noted down in a special input box.
- Conclusion – in this phase the conclusion was to be developed, based on their hypotheses and observations in the online lab. Students were shown the hypotheses that they had created before the experiment and were asked to accept or reject them based on their observations, and explain their choice.
- Discussion –this phase was intended for having students reflect on their created learning products and discuss their experience. In our study, students were asked to give feedback on two pre-defined concept maps. This process was supported by a specially developed feedback tool (see later). After this activity, students were invited to rework their own concept map (created in the conceptualization phase) and make any necessary changes. To do so students sis not have to go back to the Conceptualization phase, instead their original concept maps could be downloaded in this phase, which made the process more natural.

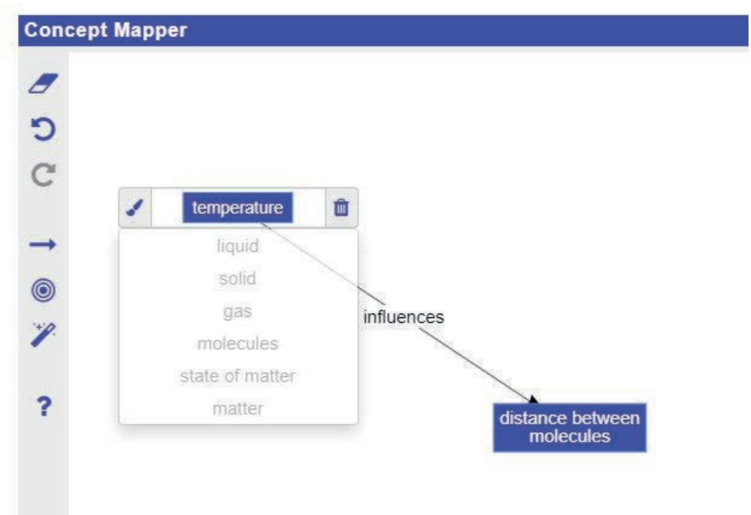

Figure 4.1. View of the Concept Mapper tool (translated into English)

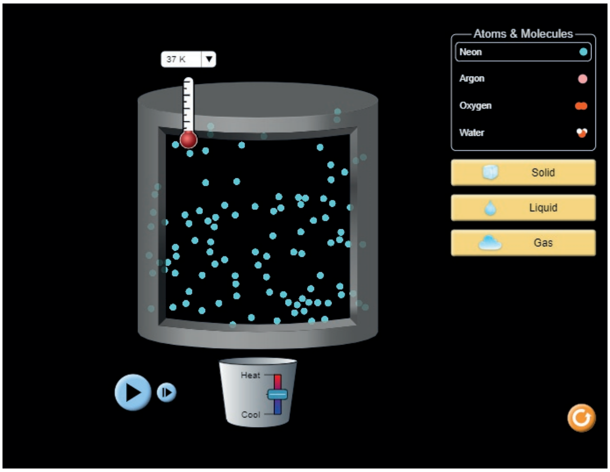

Figure 4.2. View of the online lab (English version, same as the Dutch version used in the lesson). Images by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY 4.0 https://phet.colorado.edu

Feedback tool

Students were invited to give feedback anonymously on two concept maps that were created by the research team. To give feedback, a special peer assessment tool was available. This tool presented students with a concept map by a (fictitious) peer student and a set of assessment criteria (Figure 4.3). Assessment criteria were formulated in question form (see Table 4.1) and students were supposed to give their feedback by answering these questions. Students were encouraged to give thoughtful feedback that could help these peers to improve their work. They were told that the concept maps had been created by some other students, not necessarily from the same class.

Concept maps

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

For example, one of the higher quality concept maps is presented in Figure 4.3. This concept map lacked the concept of "gas" as one of the phases and the misconception was that molecules change when going to another state of matter.

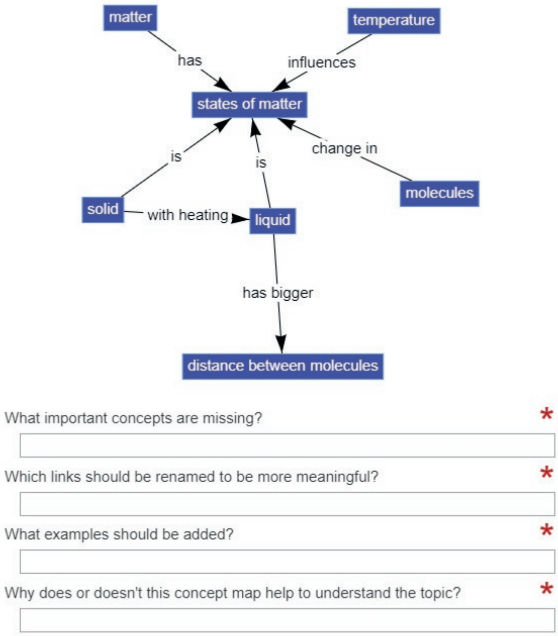

Figure 4.3. View of the feedback tool with a higher quality concept map (translated into English)

Domain knowledge test

Parallel tests were used for pre- and post-testing. The questions focused on understanding of phase-transformation processes. Each test consisted of six open-ended questions, with the maximum total score of 13 points: five questions worth two points and one question worth three points; the number of points represented the number of expected elements in the answer. Examples of questions and the grading scheme are presented in Table 4.2.

Table 4.2
Example of the grading scheme for test answers

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

Procedure

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

The content of the lesson with the ILS was part of the standard school curriculum and the lesson therefore was delivered during a regular physics lesson. A brief introduction about the procedure for the lesson was given at the beginning.

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

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

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

Analysis

To answer the research questions, pre- and post-tests were scored. A subset (41% of the pre-tests and 28% of the post-tests) were graded independently by a second rater, with Cohen's kappa being .83 and .80, respectively. Tests were scored using a scheme covering what the answer should include to obtain the points. The scheme was developed by the researchers and approved by teachers at the participating school.

Students' final versions of the concept map (after interaction with the learning environment, providing feedback, and optional revision of the concept map) were also assessed by the researchers. The following characteristics of the concept maps were evaluated:

- proposition accuracy score – the number of correct links,
- salience score – the proportion of correct links out of all links,
- complexity score – the level of complexity.

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

The salience score had a maximum of one.

The complexity scoring used a nominal scale, with the goal of discriminating between a linear representation ("sun" or "snake" shapes) and a more complex structure with more than one hierarchy level, assigning 1 point for a linear structure and 2 points for a more complex structure.

A subset (19%) of the concept maps were coded by a second rater, with Cohen's kappa of .70.

The final product that was assessed was the quality of the feedback given. Correct answers for the criteria questions were worth 1 point each, while each additional meaningful suggestion for a criteria gave them 0.5 point. A subset (25%) of the feedback was coded by a second rater, with Cohen's kappa reaching .71.

RESULTS

During the preliminary analysis of the data one participant was found with a difference between post-test and pre-test scores of more than 2.5 *SD* from the mean for the whole population. As this participant obviously did not represent the population, we labeled this participant as an outlier and this participant's data were excluded from the analysis.

The difference in pre-test scores between the conditions was not significant: $M_1 = 3.52$ ($SD = 1.60$), $M_2 = 3.91$ ($SD = 1.73$), $M_3 = 3.80$ ($SD = 2.06$), $F(2,74) = 0.34$, $p = .71$.

To check whether students learned at all, a paired t-test was conducted comparing pre- and post-test results. Post-test results were significantly higher than pre-test results with relatively large effect sizes for the whole sample [$t(76) = 4.62$, $p < .01$, Cohen's $d = 0.60$] and for each condition [$t_1(28) = 3.94$, $p < .01$, Cohen's $d_1 = 0.73$; $t_2(22) = 2.85$, $p < .01$, Cohen's $d_2 = 0.59$; $t_3(24) = 2.38$, $p < .05$, Cohen's $d_3 = 0.48$].

Post-test scores

The first outcome we present concerns the effect of condition and prior knowledge on students' test results. The descriptive statistics for participants' test results are presented in Table 4.3.

To answer the question about the differential effect of the intervention for different prior knowledge groups, students from the whole sample were divided into three groups: low prior knowledge (pre-test score more than 1*SD* below the mean), average prior knowledge (the pre-test score is within 1*SD* above or below the mean) and high prior knowledge (pre-test score more than 1*SD* above the mean).

Table 4.3
Test scores by knowledge level and condition (maximum score is 13)

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

An ANOVA was used to answer the research questions about the effect of the reviewed products' quality on the reviewers' post-test scores, with post-test score as a dependent variable and prior knowledge level and condition as independent variables. Only the main effect for prior knowledge was found to be significant [$F(2, 68) = 9.51$, $p < .01$, $\eta_p^2 = .22$], showing that higher prior knowledge led to higher post-test scores.

Inspecting the follow-up analysis showed that results for condition 1 (lower quality concept maps) could differ from those for the two other conditions, especially for the low prior knowledge group. The trends can be seen in Figure 4.4.

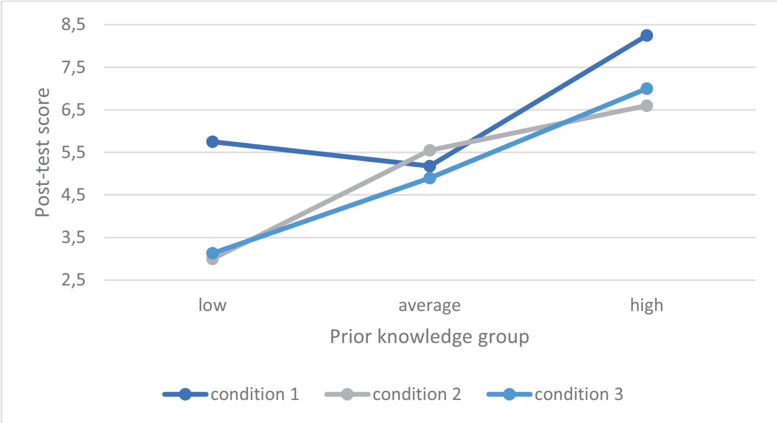

Figure 4.4. Marginal means of post-test scores (estimated) per prior knowledge group and condition

A pairwise comparison of the estimated marginal means of post-test scores between the conditions showed a statistically significant difference between condition 1 and condition 3 ($p = .048$). Students from condition 1 (lower quality concept maps) were found to have higher estimated post-test scores than students from condition 3 (higher quality concept maps) [$M_1 = 6.39$, $SE = 0.50$; $M_3 = 5.01$, $SE = 0.47$].

Concept map and feedback quality

To get a deeper understanding of the reviewers' learning process, their own learning products were assessed. A regression analysis was conducted, focusing on the role that the characteristics of the concept maps and the feedback that was given play in explaining the post-test scores. The following variables were used:

- the fact of changing one's own concept map after giving feedback,
- the number of correct propositions (proposition accuracy score),
- the proportion of correct propositions (salience score),
- the complexity of the concept map (complexity score), and
- the quality of the feedback given.

A forward regression was used to see which variables predicted the post-test score. Changing the student's own concept map was found to be a significant predictor indicating that students who had changed their concept maps scored higher on the post-test ($B = 2.01$, $\beta = .34$, $p = .005$).

As the focus of attention was on the characteristics of the learning products, prior knowledge was not included in this analysis. However, adding it to the forward regression gave similar results: both prior knowledge ($B = 1.35$, $\beta = .38$, $p = .001$) and changing concept maps ($B = 1.64$, $\beta = .27$, $p = .013$) were found to be significant predictors of post-test scores.

A MANOVA was conducted to check if the characteristics of students' own concept maps and the feedback given differed per condition. The descriptive statistics are presented in Table 4.4.

Table 4.4
Scores for concept maps and feedback quality by condition

	Condition 1 (low-quality CMs)		**Condition 2 (mixed CMs)**		**Condition 3 (high-quality CMs)**	
	$n = 26$		$n = 23$		$n = 20$[2]	
	M	*SD*	*M*	*SD*	*M*	*SD*
Changed CM after giving feedback	.23	.43	.17	.39	.30	.47
Proposition accuracy of CM	1.85	2.33	2.13	2.38	2.80	3.11
Salience score of CM	.40	.45	.51	.49	.46	.47
Complexity of CM	0.88	0.59	0.70	0.47	0.90	0.55
Feedback quality	2.43	1.07	1.82	0.90	2.03	0.98

Note. CM = concept map.

The results for the concept map characteristics showed no significant differences. As for the feedback quality, the feedback with the highest score was given by the students who were evaluating the lowest quality set of concept maps (condition 1). Students who were reviewing the mixed set of concept maps (condition 2) gave the feedback of the lowest quality. Moreover, the pairwise comparison showed that the difference in the quality of feedback between these two conditions – a lower quality set of concept maps (condition 1) and a mixed set of concept maps (condition 2) – was significant, with students from condition 1 providing higher quality feedback than students from condition 2 [$M_1 = 2.43$, $M_2 = 1.82$, $p = .033$].

[2] Due to technical reasons, the feedback provided by eight students was not saved, so their data were not included in this analysis.

Conclusion and discussion

The focus of this study was on the roles that the quality of the products that were reviewed and, in relation to that, the prior knowledge of students, play in the learning of a peer reviewer. Based on the literature (e.g., Alqassab et al., 2018a; Tsivitanidou et al., 2018), it was expected that reviewing higher quality products would be more beneficial for learning, as it would set a good example. Another expectation was that students with higher prior knowledge would benefit more from the reviewing process, as their level of knowledge would allow them to provide more criticism, so being more cognitively involved (e.g., Cho & Cho 2011; Patchan & Schunn, 2015). The findings, however, showed opposite trends.

The first finding revealed that although there was no main effect for the different quality of the products reviewed (condition), the pairwise comparison of the post-test scores between the conditions showed a statistically significant difference. According to the results, giving feedback on the set of lower quality concept maps led to higher post-test scores for the reviewers than giving feedback on the set of higher quality concept maps.

Although the results are opposite from what was expected, they can still be explained by theory. One of the explanations might be connected with the participants' prior knowledge level. Most researchers have argued that students should be competent enough in the domain to give meaningful and useful feedback, which, in turn, can lead to their own learning (see, e.g., Alqassab, Strijbos, & Ufer, 2018b). In our case, students might not have been very knowledgeable about the topic even after the lesson, as the mean post-test score was just about 5 out of 13 points. This could mean that lower quality concept maps matched their knowledge level better, so they could give meaningful feedback on this level of products. According to the trends, reviewing lower quality concept maps was especially beneficial for the lower prior knowledge group of students, which backs our argument.

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

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

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

Another explanation of the difference between the expected and obtained results could be the differences in the reviewed products. All of the studies used for the literature review invited students to assess bigger and more complex learning products – from essays and reports to geometry proofs, while the current study focused on a very compact product – a concept map. On the one hand, a concept map is more structured and, thus, the mistakes with regard to missing parts can be more obvious, which makes a contribution from weaker students more possible as they are also able to identify these mistakes. On the other hand, creating a concept map as an exercise requires much more abstract thinking, which is generally more difficult for students. Moreover, approaches to concept mapping can be quite different for different students, and understanding a different rationale used in another's concept map can also be rather difficult. Therefore, spotting mistakes in higher quality concept maps could be too challenging even for students with higher prior knowledge. These two factors together made concept maps of lower quality more preferable for students with different prior knowledge levels.

Our study was conducted in a specific context (inquiry learning) and with quite specific small products (concept maps), both of which were different form the most studies referred to. This could influence the generalizability of the findings and interpretation. However, several more general conclusions can still be drawn. First, students' prior knowledge should be taken into account while grouping them for giving feedback. Second, when dealing with smaller products, it might be more beneficial for students' learning to give them pieces of work to review that match their current level or are (a bit) lower. This would help them to associate themselves with those products and feel that they have enough knowledge to provide meaningful feedback on them. It could lead to more learning than giving them a higher quality product that they would not be able to analyze and learn from. It looks like weaker learning products stimulate students to provide more feedback that is critical and, as a result, learn more. Third, students who had reworked their own concept maps scored higher at the post-test. This means that making another step after giving feedback, namely, revisiting one's own product, could contribute to learning. Therefore, students should be encouraged to revise their own products after providing feedback. Finally, to establish generalizability of our results, future research could check whether these findings hold up for other small learning products created in the inquiry learning process, for example, hypotheses or conclusions.

References

Alqassab, M. (2017). *Peer feedback provision and mathematical proofs: Role of domain knowledge, beliefs, perceptions, epistemic emotions, and peer feedback content* [unpublished doctoral dissertation]. Ludwig-Maximilians University, München, Germany.

Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018a). The impact of peer solution quality on peer-feedback provision on geometry proofs: Evidence from eye-movement analysis. *Learning and Instruction*, 58, 182-192. doi: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. doi:10.1007/s10212-017-0342-0

Brill, J. M. (2016). Investigating peer review as a systemic pedagogy for developing the design knowledge, skills, and dispositions of novice instructional design students. *Educational Technology Research and Development*, 64, 681-705. doi:10.1007/s11423-015-9421-6

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39, 629-643. doi:10.1007/s11251-010-9146-1

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., . . . Zacharia, Z. C. (in press). Understanding teacher design practices for digital inquiry-based science learning: The case of Go-Lab. *Educational Technology Research & Development*.

Ion, G., Sánchez-Martí, A., & Agud-Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, 44, 124-138. doi:10.1080/02602938.2018.1484881

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*, 193-211. doi:10.1080/02602938.2019.1620679

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*, pp. 257-275. doi:10.1007/s11251-011-9177-2

Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2014). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, 39, 657-677. doi:10.1080/02602938.2013.860421

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, 41, 381-405. doi:10.1007/s11251-012-9236-3

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science*, 43, 591-614. doi: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. doi:10.1016/j.edurev.2015.02.003

Sluijsmans, D. M. A. (2002). *Student involvement in assessment. The training of peer assessment skills* [unpublished doctoral dissertation]. Open University of the Netherlands, The Netherlands.

Tenório, T., Bittencourt, I. I., Isotani, S., & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior*, 64, 94-107. doi:10.1016/j.chb.2016.06.020

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. *European Journal of Psychology of Education*, 33, 51-73. doi:10.1007/s10212-017-0341-1

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*, 506-519. doi:10.1016/j.learninstruc.2010.08.002

van Dijk, A. M., & Lazonder, A. W. (2013). Scaffolding students' use of learner-generated content in a technology-enhanced inquiry learning environment. *Interactive Learning Environments*, 24, 194-204. doi:10.1080/10494820.2013.834828

van Zundert, M. J., Sluijsmans, D. M. A., Könings, K. D., & van Merriënboer, J. J. G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology*, 32, 127-145. doi:10.1080/01443410.2011.626122

Wu, Y., & Schunn, C. D. (2020). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal*, 1-35. doi:10.3102/0002831220945266

Yalch, M. M., Vitale, E. M., & Ford, J. K. (2019). Benefits of peer review on students' writing. *Psychology Learning & Teaching*, 18, 317-325. doi:10.1177/1475725719835070

5

Type of reviewed products

A study of the effect of the different types of the reviewed products

This chapter is based on:

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (submitted). *Does learning from giving feedback depend on the product being reviewed: Concept maps or answers to test questions?*

Abstract

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

Introduction

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

Peer feedback

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

A number of studies have shown that the quality of the product affects how much a feedback provider (that is, a reviewer) learns (Alqassab, Strijbos, & Ufer, 2018; Patchan & Schunn, 2015; Tsivitanidou, Constantinou, Labudde, Rönnebeck, & Ropohl, 2018). One explanation for such an influence is that the quality of the product determines the type and amount of feedback. For instance, feedback on a lower quality product would involve identifying more problems and suggesting more solutions than giving feedback on a higher quality product, while spotting one mistake in a product of generally good quality could be more challenging than finding many mistakes in a lower quality product. In other words, different levels of product quality require different thinking when formulating feedback.

There can also be an interaction effect between the reviewed product's level of quality and the reviewer's prior knowledge. Because students with higher prior knowledge understand the topic and potentially know the product better, they might be better able to find mistakes and suggest solutions for improving reviewed products than students with lower prior knowledge. This difference in the feedback provided by students of different prior knowledge levels can be more obvious for low-quality products than for high-quality products, as low-quality products include more mistakes and require more improvement. For example, according to a study conducted by Patchan and Schunn (2015), students with higher prior knowledge found more mistakes and provided more suggestions for a low-quality product than for a high-quality product, while lower prior knowledge students did not distinguish between the different quality levels of products by providing different amounts of critique. This means that the way students interact with a product that they are reviewing may depend not only on the quality of the product, but also on their own prior knowledge.

These considerations are only important if the quality of the feedback they give influences the reviewers' learning, which was shown in a study by L. Li, Liu, and Steckelberg (2010). In their work, undergraduate students gave feedback on peers' project reports. The results showed that the quality of the provided feedback had a significant relationship with reviewers' own learning products. Providing better and more critical feedback facilitated students' learning, which was reflected in the quality of their own reports.

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

Type of products reviewed

The current study aims to investigate the role of the type of reviewed product in the reviewer's learning by comparing the process and results of giving feedback for two types of products: concept maps and answers to open-ended test questions. Both products were of a smaller scale compared to papers or essays. Concept mapping is a way of presenting a domain topic through concepts and connections between them. Concept maps contribute to deeper understanding of a topic, as they encourage students to think about such connections (e.g., Novak, 2010). Schroeder, Nesbit, Anguiano, and Adesope (2018) conducted a meta-analysis on the effect of concept mapping on learning, which included 63 studies with 142 independent effect sizes. The authors demonstrated a moderate overall effect ($g = 0.58$, $p < .001$) of concept mapping on learning. Moreover, this effect was found across different domains, with a nonsignificant difference between STEM (science, engineering, technology and math) and non-STEM domains. The same meta-analysis found that working with concept maps was especially beneficial for secondary school students ($g = 1.24$, $p < .001$).

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

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

Research questions

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

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

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

Method

Participants

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

In each class, students were randomly assigned to one of the two conditions: giving feedback on concept maps (CM condition) and giving feedback on answers to open-ended test questions (test condition). The final sample of 127 participants included 66 students in the concept map condition and 61 students in the test condition.

Due to technical reasons, the feedback provided by two students (one in each condition) was not saved, so their data were not included in the analysis of the quality of the feedback. However, their data were used for the other parts of the analysis, as they did complete both lessons and gave feedback.

Study design

In the current study, participants were asked to give feedback on (fictitious) peers' learning products. In both conditions, they were supported in this process by assessment criteria that were relevant to the product type. In one condition, students were asked to give feedback on two concept maps. They were assisted in this process by pre-defined assessment criteria. These criteria were presented in a question form and guided students in the direction of the desired features of a concept map. The assessment criteria were rather general and not specific for a domain. This approach was based on the study by van Dijk and Lazonder (2013), in which the most important characteristics of a concept map were discussed. The questions used in this study as guidelines for giving feedback on concept maps are presented in Table 5.1.

Table 5.1
Assessment criteria for giving feedback on concept maps (translated from Dutch)

1.	What important concepts are missing?
2.	How would you change the structure of the map?
3.	Which links should be renamed to be more meaningful?
4.	What examples should be added?
5.	Why is this concept map helpful or not helpful for understanding the topic?

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

All reviewed learning products (concept maps and test answers) were created by the research team. This was done for experimental control to ensure that participants assessed the same products and had the same opportunities for learning from giving feedback on them. To create a more realistic setting, students were told that the products came from other students, probably from a different school.

Materials

The lessons were given in the students' native language (Dutch) and followed the national curriculum for chemistry. The topic fit into the regular programme of study, and presented the theme of matter and elements, and in particular, atomic structure.

The lessons were built with the help of the Go-Lab ecosystem (see www.golabz.eu). This ecosystem allows the creation of inquiry learning spaces (ILSs) that guide students through the inquiry process (de Jong et al., in press). By following an inquiry learning cycle, students explore a scientific phenomenon in a way resembling scientific research. The cycle includes several phases (orientation, conceptualization, investigation, conclusion and discussion), each having a particular purpose in the inquiry process (Pedaste et al., 2015).

The ILS created for the current study followed an adapted basic scenario for inquiry learning and consisted of the following phases (the names are translated from the Dutch):

- Starting – the topic of the lesson, its goal and procedure were introduced to the students. They were asked to write down their ideas about atomic structure as a starting point for their investigation. A brief informational video giving the main concepts of the topic was presented to ensure that all students had the same starting point in studying the topic.
- Preparing - the idea of concept mapping as a way to present a topic was introduced to students by an instructional video and written instructions. Students were invited to create a concept map representing the topic of atomic structure using a specially designed tool – the Concept Mapper (see Figure 5.1). The tool served not only as a canvas, but also as a scaffold, at it included some pre-defined relevant concepts and link names. However, students could also type in their own concepts and links.
- Investigating – the exploration of the atomic structure was done with the help of an online lab (see Figure 5.2). In the lab, students could build an atom out of the components, see which element this atom represents and how it can change into an ion (a charged particle). Students were guided through the investigation process by a series of tasks, but they could complete them in any order or come back to them. At the end of the phase, students were asked again to describe the atomic structure as they could do it at that moment.
- Refreshing – this phase was added, as the inquiry process took two lessons and some connection with the previous lesson needed to be established. Students were shown their conclusion from the previous lesson, and brief information and a small quiz were used to refresh their knowledge about the topic.
- Applying – understanding of the topic was checked and reinforced with a help of game based on the material that had been studied. Students needed to apply what they had learned to complete the tasks, for example, identifying an element by its components or building a particle with a given mass and charge.
- Assessing – giving feedback on what students were told were peers' products (concept maps or answers to test questions) was done in a special peer assessment tool. After giving feedback, all students (regardless of the condition they worked in) were given an opportunity to re-work their own concept maps created in the Preparation phase in the previous lesson. They could download their original concept maps and make any necessary changes.

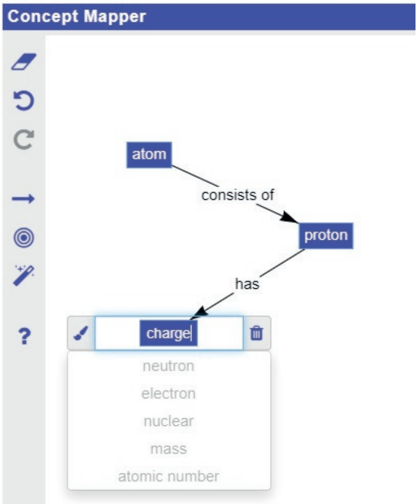

Figure 5.1. View of the Concept Mapper tool (translated from Dutch)

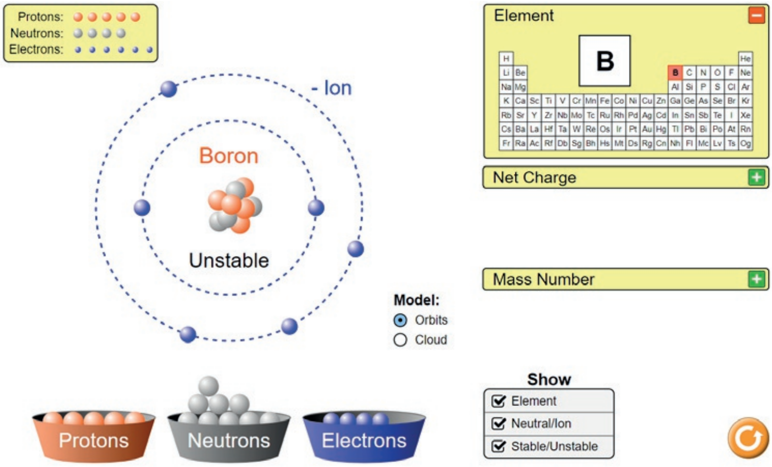

Figure 5.2. View of the online lab (English version, same as the Dutch version used in the lesson). Images by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY 4.0 https://phet.colorado.edu

All "peer" products that students gave feedback on were created by the research team. This was done to provide everyone with the same quality of products to assess, as the quality of the reviewed product influences the feedback-giving process and its outcomes (see, e.g., Patchan & Schunn, 2015). The products for both conditions were of medium quality, presenting both right and wrong information. Moreover, products for both conditions included the same three misconceptions: an electron has a mass, a neutron has a charge, and neutrons define the element.

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

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

Students gave their feedback anonymously using a special peer assessment tool. The tool allowed them to see the reviewed products and the assessment criteria and to write their comments. The view of the peer assessment tool for both conditions is presented in Figures 5.3a and 5.3b.

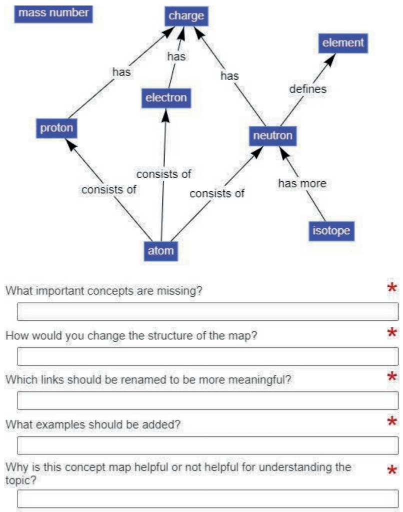

Figure 5.3a. View of the peer assessment tool for the CM condition (translated from Dutch)

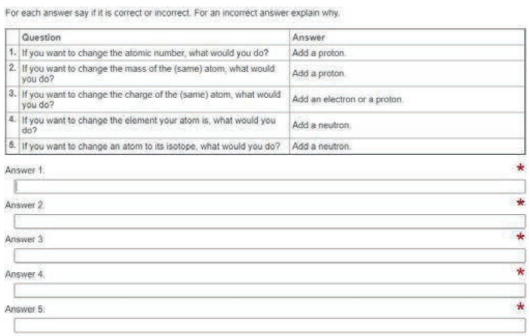

Figure 5.3b. View of the peer assessment tool for the test condition (translated from Dutch)

Pre-tests and post-tests were parallel versions, and consisted of six open-ended questions each. The tests were considered parallel, as they addressed the same concepts but in differently formulated questions. The maximum score for each test was 11 points, with points per question varying between one and three. There were three types of questions used: those checking knowledge (1 point), those requiring application of knowledge (2 points), and those requiring understanding the connections between the concepts (3 points). Examples of each type of question are given in Table 5.2. Moreover, for questions with more than one point, a coding scheme was used for a partially correct answer. It awarded more points for argumentation explaining the connection between the concepts (see Table 5.3 for an example).

Table 5.2
Examples of the test questions (translated from Dutch)

Question	Points
What does the atomic number give information about?	1
A particle has 6 protons, 5 neutrons and 7 electrons. What is the charge of the particle?	2
A particle has 16 electrons. Can you say what element it is? Why/why not?	3

Table 5.3
Example of a coding scheme for a test answer (translated from Dutch)

Question	Answers	Points
A particle has 16 electrons. Can you say what element it is? Why/why not?	No, I cannot.	1
	No, I cannot. We need to know the number of protons.	2
	No, I cannot. We need to know the number of protons, because protons define what element it is.	3

Procedure

The experiment lasted two school lessons (50 minutes each) and took place during two successive chemistry lessons, according to the regular timetable of the participating classes. The two experimental lessons occurred within five days for all classes. Both tests (pre- and post-) and working in the ILS occurred online, with each participant working individually on a device.

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

The first lesson included the pre-test and the first three phases of the ILS (starting, preparing and investigating). Students were given 15 minutes for the pre-test, but they could proceed to the ILS if they were finished with the pre-test early. Students could move freely through the first three phases of the ILS, but could not go further.

The second lesson included the last three phases of the ILS (refreshing, applying, and assessing) and the post-test. Students could now move freely between all phases of the ILS. Students were given 15 minutes for the post-test, but they could proceed to it earlier if they finished working in the ILS; in that case, they could not return to the

ILS. Before students could go on to the post-test, the researcher checked that they had given the requested feedback, and reminded them to do so if necessary.

Analysis

First, the pre-test and post-test consisted of open-ended questions that needed to be scored. The scoring was done applying the coding scheme, which was developed by the researchers and approved by participating teachers. The scheme included the right answers and corresponding points. A second rater graded 20% of the pre-tests and post-tests to estimate the inter-rater reliability, which was found to be good, with Cohen's kappa being .92 for the pre-test and .74 for the post-test.

Second, the final version of a student's own concept map (after giving feedback) was coded. The fact of students' changing their concept maps after giving feedback was coded as one, while not changing as zero. The following criteria were used to code characteristics of a concept map:

- proposition accuracy score – the number of correct links;
- salience score – the proportion of correct links out of all links used;
- complexity score – the level of complexity.

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

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

A second rater coded 20% of the concept maps, with Cohen's kappa reaching .62. This is an acceptable yet moderate result; however, the scales for assessed characteristics, were continuous scales. Cohen's kappa is less informative for a continuous scale variable, so a Pearson's correlation was also used to check the inter-rater agreement, with $r = .95$ ($p < .01$). Together with the Cohen's kappa value, this suggested that the scoring was reliable.

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

For the test condition, the aim of the feedback was to identify incorrect answers and provide an explanation of what was wrong. One point was given for each accurate identification of the correctness/incorrectness of an answer and one point for each viable explanation of why an answer was incorrect. With three incorrect answers out of five, the maximum score for the quality of feedback was eight points. A second rater coded 20% of the feedback given for both conditions, with Cohen's kappa being .86.

Results

The difference in pre-test scores between the conditions was not significant: M_{CM} = 2.00 (SD = 1.47) and M_{TEST} = 1.97 (SD = 1.73), $t(125) = 0.12$, $p = .91$ (see Table 5.4).

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

The analysis covered the post-test scores, the quality of students' own concept maps and the quality of the feedback given. The results are presented below.

Learning gain

As a prerequisite for the further analyses, it was first checked whether students learned during the experimental lessons. Paired samples *t*-tests were used for pre-test and post-test scores. The results showed that students in both conditions learned during the lessons [$t_{CM}(65) = 11.91$, $p < .01$, Cohen's $d = 1.47$; $t_{TEST}(60) = 10.92$, $p < .01$, Cohen's $d = 1.40$; $t(126) = 16.21$, $p < .01$, Cohen's $d = 1.44$].

Descriptive statistics for all prior knowledge groups and both conditions are presented in Table 5.4.

Table 5.4
Test scores by knowledge level and condition (maximum score is 11)

	Low prior knowledge		**Average prior knowledge**		**High prior knowledge**		**Total**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CM condition	*n* = 4		*n* = 51		*n* = 11		*n* = 66	
pre-test	0.00	0.00	1.57	0.70	4.78	0.91	2.00	1.47
post-test	5.00	4.00	6.73	2.58	6.18	2.93	6.53	2.71
Test condition	*n* = 10		*n* = 39		*n* = 12		*n* = 61	
pre-test	0.00	0.00	1.56	0.72	4.92	1.00	1.97	1.73
post-test	5.80	3.23	6.00	3.00	7.42	2.68	6.25	3.02
Total	*n* = 14		*n* = 90		*n* = 23		*n* = 127	
pre-test	0.00	0.00	1.57	0.70	4.83	0.94	1.98	1.59
post-test	5.57	3.32	6.41	2.80	6.83	2.81	6.39	2.85

Effect of condition on post-test scores

Second, an ANOVA was conducted to answer the research questions about the influence of the type of product reviewed (concept maps or answers to test questions) and prior knowledge on the post-test scores. Condition and prior knowledge level were used as independent variables and post-test sore as a dependent variable. No main effect was found to be statistically significant for either of the variables, nor was there an interaction effect. None of the pairwise comparisons was significant either.

The quality of students' own concept maps

As students in one condition reviewed concept maps and students in the other condition reviewed answers to test questions, but students in both conditions had to create their own concept maps, we expected there to be a difference between the conditions in the quality of the final concept maps produced, which was checked with a *t*-test. Apart from the coded parameters for the quality of a concept map: the proposition accuracy (the number of correct links), the salience score (the proportion of the correct links) and the complexity of the concept map, the fact of students' changing their own concept maps or not after giving feedback was also included in the analysis. No statistically significant differences were found for any of the parameters. The descriptive statistics for these variables are presented in Table 5.5.

Table 5.5
Concept map characteristics by condition (maximum scores' explanation is given in the Analysis part)

	Maximum score	**CM condition**		**Test condition**	
		n = 66		*n* = 61	
		M	*SD*	*M*	*SD*
Changed CM after giving feedback	1	.62	.49	.66	.48
CM proposition accuracy	n/a	7.30	3.60	7.11	4.02
CM salience	1	.66	.26	.62	.30
CM complexity	2	1.92	0.32	1.82	0.50

Note. CM = concept map.

A regression analysis including the quality of students' own concept maps (the final versions after giving feedback), the fact of their changing their concept maps after giving feedback and the quality of the feedback given was conducted, to see if these factors predicted post-test scores. The quality of the feedback given and the proposition accuracy (number of correct links) were found to be significant predictors of the post-test score, with the post-test score increasing by 0.34 points when the proposition accuracy increased by 1 point and all other variables stayed the same. The regression coefficients are shown in Table 5.6.

Table 5.6
Regression coefficients for quality of concept map and feedback given to predict post-test scores

	B	***t***	***p***
Changed CM after giving feedback	0.603	1.281	.203
CM proposition accuracy	0.340	4.011	.000
CM salience	-0.960	-0.877	.382
CM complexity	0.285	0.486	.628
Quality of feedback given	0.479	3.402	.001

Note. CM = concept map.

The quality of feedback given

Even though the products to give feedback on were different in the two conditions, students in both conditions were supported in the process of giving feedback by assessment criteria matching the product type.

A *t*-test showed that the difference in feedback quality was significant [$t(123) = -2.37$, $p = .019$], with students giving feedback on answers to test questions providing feedback of higher quality than the students giving feedback on concept maps (M_{CM}

= 2.53, *SD* = 1.14, M_{TEST} = 3.18, *SD* = 1.89). (Due to technical reasons, the feedback from two participants was not saved, so the feedback of 125 participants was used for this analysis.)

As the regression analysis above demonstrated, the quality of the feedback given was a significant predictor of post-test scores. A separate regression analysis was conducted with the quality of feedback and prior knowledge predicting the post-test scores. Prior knowledge was not found to be a significant predictor, while the quality of the feedback given was [$F(2, 122) = 7.95$, $p < .01$, $R^2 = .12$]. The coefficients for the regression are shown in Table 5.7.

Table 5.7
Regression coefficients for the quality of feedback given and prior knowledge to predict post-test score

	B	***t***	***p***
Quality of feedback given	0.569	3.663	.000
Prior knowledge	0.168	1.097	.275

When analysed separately, the results for conditions were similar. For the condition giving feedback on concept maps [$F_{CM}(2, 62) = 8.82$, $p < .01$, $R^2 = .22$], an increase of 1 point in the feedback quality would lead to the increase of 1.13 points in the post-test score. For the condition giving feedback on answers to test questions [$F_{TEST}(2, 57) = 3.70$, $p < .05$, $R^2 = .12$], an increase of 1 point in the feedback quality would lead to the increase of 0.37 points in the post-test score.

To further investigate the connection between the prior knowledge level and the quality of feedback, and since no correlation was found between prior knowledge and post-test scores, an additional correlation analysis was conducted. Its purpose was to check whether prior knowledge and the quality of feedback were connected. The correlation was not found to be statistically significant ($p = .158$).

Conclusion and discussion

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

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

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

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

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

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

Finally, and probably the most importantly, the quality of the feedback provided was found to predict post-test scores: the higher the quality of the provided feedback, the higher the post-test score. An indication that feedback quality could predict post-test scores was also found in one of our previous studies (Study 3). At that point, there was not enough evidence to say that better feedback led to a higher post-test score, as they both could have been caused by higher prior knowledge. However, in the current study, prior knowledge was not found to be a significant predictor of post-test scores. Moreover, the quality of the feedback given was not found to be correlated with prior knowledge. This may mean that the quality of feedback indeed explained the post-test score and if students invested more effort into giving better feedback, it could lead to more learning for reviewers themselves. This was even more the case for students in the concept map condition, in which giving good feedback was probably more challenging than in the other condition, thus, requiring more cognitive involvement. It might seem contradictory to find significant differences in the quality of feedback between the conditions (and the fact that the quality of feedback could predict post-test scores) and not to find a significant difference in post-test scores. A possible explanation is that the variability in post-test score was too high to reveal an effect of condition. Moreover, quality of feedback explained about 12% of the post-test scores, indicating the presence of other explanatory factors. The interpretation of the correlation between the quality of provided feedback and the post-test scores for practice may mean that the effect from giving feedback does not depend that much on the product type, as long as students provide meaningful comments.

The finding that feedback quality can play an important role in learning was also supported by other studies. For example, the study by L. Li et al. (2010) demonstrated a significant relationship between the quality of provided feedback and the quality of students' own projects. The authors suggested as the ability to give meaningful and constructive feedback is very important for a reviewer, time and effort should be invested in developing this ability. Therefore, an interesting direction for further research would be studying what factors influence such an ability and identifying ways to increase the quality of peer feedback. The research done so far has pointed in the direction of structuring the process of giving peer feedback. For example, Gielen and De Wever (2015) found that peer feedback quality was higher in the group with an elaborated feedback structure than in the group with no provided structure for giving feedback.

In conclusion, this study investigated providing peer feedback on rather small-scale products created in an inquiry learning environment. Several take-away messages emerged from this study. First, the type of reviewed product does not seem to play an important role in the reviewers' learning as long as they provide good quality feedback. Second, even with a rather brief intervention (which is caused by small-scale products), the quality of peer feedback explained post-test scores, which demonstrates the value of feedback-providing activities for learning. Further research studying peer feedback will be beneficial for educational practice.

References

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

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

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

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

de Jong, T., Gillet, D., Rodríguez-Triana, M. J., Hovardas, T., Dikke, D., Doran, R., . . . Zacharia, Z. C. (in press). Understanding teacher design practices for digital inquiry-based science learning: The case of Go-Lab. *Educational Technology Research & Development*.

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning*, *31*, 435-449. doi:10.1111/jcal.12096

Ion, G., Sánchez-Martí, A., & Agud-Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, *44*, 124-138. doi:10.1080/02602938.2018.1484881

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

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*, 193-211. doi:10.1080/02602938.2019.1620679

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

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

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

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

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science*, *43*, 591-614. doi: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. doi: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. doi:10.2308/iace-50754

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

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. *European Journal of Psychology of Education*, *33*, 51-73. doi:10.1007/s10212-017-0341-1

van Dijk, A. M., & Lazonder, A. W. (2013). Scaffolding students' use of learner-generated content in a technology-enhanced inquiry learning environment. *Interactive Learning Environments*, *24*, 194-204. doi:10.1080/10494820.2013.834828

Wu, Y., & Schunn, C. D. (2020). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal*, 1-35. doi:10.3102/0002831220945266

Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*, 372-386. doi:10.1080/02602938.2019.1644603

6

General discussion

INTRODUCTION

Traditionally, feedback from peers was used when teachers could not provide proper feedback themselves, for example, because of large class sizes (e.g., Falchikov, & Goldfinch, 2000). This way of using peer assessment was not adopted by all teachers, due to its low reliability and validity (e.g., Liu & Carless, 2006). However, there has been a shift in goals, from using peer feedback as a replacement for teachers' feedback to using it as a learning tool (e.g., Adachi, Tai, & Dawson, 2018). And as a learning tool, giving feedback to peers can be a learning experience for feedback providers (or reviewers). Various factors can influence such learning, so that knowing more about these factors will lead to better understanding of the process of giving feedback to peers and how best to organize it. Therefore, the aim of this dissertation was to investigate these factors, on the one hand, in order to contribute to scientific knowledge about the feedback-giving process and, on the other hand, to provide practitioners with recommendations for organizing feedback-giving activities in a way that is most beneficial for reviewers' learning.

This final chapter presents a summary of the results and their meaning for educational practice, as well as suggestions for directions for future research.

Characteristics of the studies conducted

In this section, we discuss how the design of the studies that were conducted differentiates them from other studies in the field of peer feedback. These design features are the same for all of the studies conducted; even though they are described in each study, summarizing them here can provide a better context for interpreting the results.

First, the focus in all of the studies was on *giving feedback to peers*. Giving feedback is one part of the peer assessment process, whereas receiving feedback is the other part. The choice of this focus was influenced by the fact that several studies have shown that giving feedback to peers is more beneficial for students' learning than receiving feedback from peers (e.g., Li & Grion, 2019; Li, Liu, & Steckelberg, 2010; Lundstrom & Baker, 2009), while feedback giving has also been less studied than feedback receiving. Separating these two parts of the peer assessment process and focusing only on giving feedback was done in order to get a better understanding of the factors that might influence reviewers' learning. Knowing about such factors can define the best way to organize giving feedback to peers as a full learning experience for reviewers.

Second, all of our studies covered STEM topics and were conducted in a special learning context: they used online *inquiry learning* environments created in the Go-Lab ecosystem (www.golabz.eu). Inquiry learning imitates a real scientific research cycle and facilitates students' following of this cycle, which is especially valuable for science subjects in a secondary school curriculum. Adding a feedback-giving activity in an inquiry learning context makes the inquiry learning cycle even closer to the real research cycle, as giving feedback on peers' products (such as articles, presentations, proposals, etc.) is a natural part of a scientist's work. Critiquing peers' learning products and providing suggestions for their improvement allows students to develop conceptual understanding of a topic and scientific reasoning skills (e.g., Dunbar, 2000; Friesen & Scott, 2013). Moreover, giving feedback on peers' products provides students with another opportunity to reflect on and revise their own products, which may also stimulate learning. Therefore, studying the process and learning outcomes of giving feedback to peers in an inquiry learning context might lead to better understanding of the different aspects involved in giving feedback than studying it in the context of traditional instruction.

Finally, in all of the studies, the *products to be reviewed* were *relatively small-scale* and the *feedback-giving moment* was *brief*, which was an intentional design decision. As all the products that students gave feedback on were part of a larger inquiry lesson, such products were on a smaller scale compared to self-standing products such as essays, papers, web page designs, or projects. While using smaller scale products fit the context (inquiry learning), the target group (secondary school students), and the domain (STEM) better than larger scale products, it also meant that the feedback-giving intervention was rather brief. And as giving feedback on a smaller scale product took less time, it could also take less mental effort. In other words, it was not yet clear whether a rather brief intervention could stimulate enough cognitive activity by reviewers to influence their learning outcomes. Investigating the giving of feedback on smaller scale products could help with the development of practical recommendations, which, in turn, could assist with using this feedback technique in an everyday classroom context.

RESULTS OF CONDUCTED STUDIES AND THEIR MEANING FOR PRACTICE

The dissertation was inspired by the peer-feedback model developed by Sluijsmans (2002), which includes three steps: defining assessment criteria, judging a peer's performance, and providing feedback for future learning. Each study focused on one of the steps of the model and investigated this step's contribution to reviewers' learning. All studies also included reviewers' prior knowledge in the analysis, as several studies have demonstrated that students' prior knowledge can influence the process of giving feedback and its outcomes (e.g., Patchan, Hawk, Stevens, & Schunn, 2013; Tsivitanidou, Constantinou, Labudde, Rönnebeck, & Ropohl, 2018; van Zundert, Könings, Sluijsmans, & van Merriënboer, 2012). Below, the results of all of the conducted studies are summarized and discussed, following the order of the steps in the model of the feedback-giving process that was used (Sluijsmans, 2002).

The role of assessment criteria (Study 2)

As the first step in giving feedback, according to the model by Sluijsmans (2002), is defining assessment criteria, this study (Study 2) investigated the role of being provided with assessment criteria for giving feedback on peers' concept maps. In the literature, two opposite views are found concerning the effect of provided assessment criteria. One point of view is that students need more guidance in the process of providing feedback and that assessment criteria can play a role of such guidance by demonstrating important characteristics of learning products (e.g., Gan & Hattie, 2014; Gielen & De Wever, 2015; Panadero, Romero, & Strijbos, 2013). Therefore, students supported by assessment criteria should have better learning results than students who are not supported. The other point of view is that it is easier for students to use their own assessment criteria than to apply assessment criteria with which they are provided. This can be especially true for complex subjects, as provided criteria might be too difficult or abstract for students to use, while their own criteria are always at their own level (e.g., Jones & Alcock, 2014; Orsmond, Merry, & Reiling, 2000).

The existing literature presents inconclusive results on whether being provided with assessment criteria leads to more or less learning. In our study, there was no difference found in post-test scores (controlling for prior knowledge) between students who were provided with assessment criteria and those who were not. In addition, it was clear that students could give meaningful and content-related feedback even without being provided with assessment criteria. These results support the previously suggested view that secondary school students are able to provide feedback on their peers' products, and they do not need to have provided assessment criteria to do so (e.g., Tsivitanidou, Zacharia, & Hovardas, 2011).

The fact that reviewers had to think of their own assessment criteria did not lead to their lower learning performance. In a real classroom situation, not providing assessment criteria may save teachers' time otherwise spent to provide criteria and students' time otherwise spent to understand the provided criteria. This result also suggests that teachers can choose settings for giving feedback (with or without criteria) depending on the situation, but neither option would lead to less learning for reviewers. Instead, teachers can focus their attention on encouraging students to give meaningful feedback to peers.

The role of the quality and type of reviewed products (Studies 3 and 4)

The second step of the feedback-giving process consists of judging peers' performance (Sluijsmans, 2002). Research done so far has indicated that the characteristics of the reviewed products may influence the feedback provider's learning (e.g., Patchan & Schunn, 2015). Therefore, two studies were conducted to investigate the role of the reviewed product's characteristics in reviewers' learning. In one of them, students in different conditions reviewed concept maps of different levels of quality (Study 3), and in the other, students gave feedback either on concept maps or on answers to open-ended test questions (Study 4).

Judging peers' products could require reviewers to perform several cognitive activities. According to Hattie and Timperley (2007), such activities would include analysing the existing state of a product, comparing it against assessment criteria, and suggesting directions for improvement based on identified problems or mistakes. The quality of reviewed products should affect these activities: low- and high-quality products might require different analyses, the outcome of a comparison against assessment criteria would be different, as different products can have different number of mistakes and, finally, products of different levels of quality would require different recommendations. In other words, reviewed products of different levels of quality would require feedback with different content, and students would need enough expertise to provide it (e.g., Alqassab, Strijbos, & Ufer, 2018b). This means that the quality of reviewed products together with reviewers' expertise (prior knowledge) may lead to different learning outcomes for reviewers. However, students do not learn from reviewing only by identifying problems and by suggesting solutions. They can also identify and adopt successful strategies for completing the task; in other words, they can learn from good examples found in high-quality products (Alqassab, Strijbos, & Ufer, 2018a; Tsivitanidou et al., 2018). Therefore, it is not yet clear which level of quality (low or high) of reviewed products could lead to more learning for reviewers.

Similarly, different types of products may stimulate different types of feedback, which, in turn, could lead to different learning by reviewers. Giving feedback on concept maps might lead to more conceptual learning than giving feedback on answers to test questions, as concept maps may stimulate more conceptual thinking than answers to test questions (e.g., Chen & Allen, 2017). However, spotting mistakes in a more complex and less familiar product, such as a concept map, could also be more challenging than spotting mistakes in answers to test questions. Being able to identify mistakes and suggest improvements while reviewing can indicate understanding of the topic and the product, and can form a basis for reflection. Therefore, identification of fewer problem areas may indicate less understanding and lead to less learning.

Comparison of conditions that reviewed products of different levels of quality (Study 3) showed that students reviewing lower quality concept maps had higher post-test scores (controlling for prior knowledge) than students reviewing higher quality concept maps. Moreover, the quality of the feedback provided by students reviewing low quality products was also higher than in two other conditions in which students reviewed concept maps of either higher or mixed quality. Comparison of conditions in which students reviewed concept maps or answers to test questions (Study 4) did not reveal a difference in post-test scores (controlling for prior knowledge) between the conditions. However, the quality of feedback given on answers to test questions was higher than the quality of feedback given on concept maps. Moreover, the quality of the provided feedback was found to predict post-test scores for both conditions.

The results might suggest that students felt more comfortable with and, as a result, were better at giving feedback on lower quality products than on higher quality ones, as they could see more mistakes and make more suggestions, which led to better learning outcomes. This is in line with the results of Study 4, in which the quality of feedback given on answers to test questions was higher than the quality of feedback given on concept maps. As mentioned above, answers to test questions were generally a more familiar product than concept maps, so students could spot mistakes more easily and provided feedback of higher quality on answers to test questions than on concept maps. In both studies, an easier and more familiar product to review stimulated higher quality feedback.

We can conclude that the quality of feedback that students provide plays an important role in reviewers' learning. To contribute the most to this learning, the products to be reviewed should be at the same or lower level of quality than the student's current level of performance. This understanding can help teachers to choose a matching strategy for peer feedback, assigning students of approximately the same level to work together. Moreover, the type of product to review does not seem to influence learning as long as students give high-quality feedback. Investing in their doing so, for example, by providing more training on giving feedback, could be a good way to increase learning from giving feedback.

The role of feedback method (Study 1)

This study (Study 1) concerned the final step in the feedback-giving model by Sluijsmans (2002); in particular, it investigated whether the method in which peer feedback was provided influenced reviewers' learning originating from it. The study compared the learning of students in two conditions: those who gave feedback by commenting or those who gave feedback by giving grades using smileys. Providing comments on peers' products has been shown to be more beneficial for reviewers' learning than providing grades (e.g., Wooley, Was, Schunn, & Dalton, 2008, July; Xiao & Lucking, 2008). This can be attributed to the fact that students are more cognitively involved with the material while commenting than while grading, as they not only evaluate peers' products, but also provide some recommendations for improvement. However, commenting is more time- and effort-consuming than grading if the learning products are large-scale (e.g., essays, reports, projects), which is often seen as a disadvantage of using comments as a way to give feedback. Reviewing smaller scale products (such as concept maps) does not require much time, making the difference in time and effort between the two conditions very small. Therefore, comparing commenting and grading as two ways of giving feedback on smaller scale products could reveal new trends or confirm the existing ones.

The results showed that students from the commenting condition had higher post-test scores (controlling for prior knowledge) than students who graded peers' concept maps with smileys. The benefit of commenting was especially obvious for students with lower prior knowledge. Finding a differential effect of commenting for prior knowledge groups made it clear that a more targeted approach is needed when implementing peer feedback. Prior knowledge is an important factor in the feedback-giving process, as reviewers need to be competent enough to provide meaningful feedback (e.g., Alqassab et al., 2018b; van Zundert et al., 2012). The fact that even students with lower prior knowledge could learn from giving comments on peers' concept maps showed that with a manageable task level, all students could benefit from this process.

Giving feedback by commenting led to more learning for reviewers than giving feedback by grading with smileys. This result may help teachers to see the value of having students comment on their peers' small-scale products. Having a brief peer feedback event when students give comments can make it a good fit for a regular lesson, while still reaping all of the benefits for reviewers' learning from providing feedback.

Practical recommendations

The following recommendations for practice can be presented based on the obtained results.

First, from the results of Study 1, it became clear that how students gave feedback did influence the learning originating from this process. When students, especially students with low prior knowledge, are asked to give feedback on peers' work, it can more beneficial for their learning to comment than to grade.

Second, the results of Study 2 indicated that secondary school students do not need to be provided with assessment criteria to be able to give meaningful content-related feedback. This means that teachers can choose to provide assessment criteria or not without decreasing this type of learning.

Third, the results of Studies 3 and 4 demonstrated that the characteristics of the product reviewed might influence reviewers' interaction with the material and their learning. To learn the most, students should be asked to give feedback on a product at their current level of performance or lower; in addition, familiarity with the type of product can be a plus.

Fourth, Studies 2 and 4 revealed that the quality of the feedback provided plays an important role in reviewers' learning. In particular, higher quality feedback was associated with higher post-test scores. Therefore, to maximize students' learning, they should be encouraged to give high-quality feedback.

Finally, there was what could be called a *side* conclusion; side, as it was not a primary focus of the research. As the results of Studies 1 and 4 showed, the number of correct links in students' own concept maps predicted their post-test scores. In other words, concept maps reflected students' knowledge of the domain. This means that the creation of concept maps could be used in science education in secondary school as a way to demonstrate and check students' understanding of a topic.

FUTURE RESEARCH DIRECTIONS

This dissertation aimed at investigating the factors that may influence reviewers' learning from giving feedback to peers. Even though several conclusions were drawn, it also became evident that further research on the feedback-giving process and its outcomes is needed. This research can take several directions, which are introduced below.

The results of several of the studies that were conducted showed that higher quality feedback provided by students was associated with their higher learning gains. Therefore, studying how to increase the quality of feedback given by students could be helpful for practitioners. There can be several reasons for low-quality feedback. One may be that students do not have enough knowledge about the product to provide high-quality feedback on it (e.g., van Zundert et al., 2012). In our studies, students gave feedback on concept maps – a product that was quite new for them. Therefore, it would be interesting to investigate whether giving students more training in concept mapping would increase their ability to give good feedback on concept maps. Another reason can be that students' lack of general evaluative skills that are important for giving feedback (e.g., Rotsaert, Panadero, Schellens, & Raes, 2018). Therefore, it might be worth studying whether training students in giving feedback on various products would lead to their providing generally better feedback. Investigating either of these reasons can provide more understanding of how to help students give high-quality peer feedback.

Another direction for further research could involve checking whether the current findings stand when students give feedback on other products created in inquiry learning environments, for example, hypotheses or experimental designs. Investigating this can contribute to creating a bigger picture of the use of peer feedback in an inquiry learning context. As a combination of giving peer feedback and inquiry learning can enrich both of these approaches, finding an optimal way to organize giving feedback in inquiry learning may lead to its wider implementation.

CONCLUDING NOTES

When properly organized, giving feedback to peers can be a learning experience for a feedback provider even when reviewing a small-scale product. This makes giving peer feedback more applicable in a real classroom situation, as teachers do not have to change a lot in the lesson to include a feedback-giving activity on a smaller product. Using online platforms (such as Go-Lab) can make giving feedback more natural and easier than in traditional instruction, due to the ability to configure parameters of the feedback-giving process according to the learning goals. While research on this topic carries on, peer assessment should be implemented in secondary schools more often, with a view to benefiting feedback providers.

References

Adachi, C., Tai, J. H.-M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education, 43*, 294-306. doi:10.1080/02602938.2017.1339775

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. doi: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. doi:10.1007/s10212-017-0342-0

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

Dunbar, K. (2000). How scientists think in the real world: implications for science education. *Journal of Applied Developmental Psychology, 21*, 49-58. doi:10.1016/S0193-3973(99)00050-7

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*, 287-322. doi:10.3102/00346543070003287

Friesen, S., & Scott, D. (2013). *Inquiry-based learning: A review of the research literature* [Paper prepared for the Alberta Ministry of Education]. Retrieved from: https://galileo.org/focus-on-inquiry-lit-review.pdf

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*, 861-878. doi:10.1007/s11251-014-9319-4

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435-449. doi:10.1111/jcal.12096

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112. doi:10.3102/003465430298487

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*, 1774-1787. doi:10.1080/03075079.2013.821974

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

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

Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education, 11*, 279-290. doi:10.1080/13562510600680582

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. doi:10.1016/j.jslw.2008.06.002

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*, 23-38. doi:10.1080/02602930050025006

Panadero, E., Romero, M., & Strijbos, 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*, 195-203. doi:10.1016/j.stueduc.2013.10.005

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science, 41*, 381-405. doi:10.1007/s11251-012-9236-3

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science, 43*, 591-614. doi:10.1007/s11251-015-9353-x

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*, 255-275. doi:10.1007/s10212-017-0329-x

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.

Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. *European Journal of Psychology of Education, 33*, 51-73. doi:10.1007/s10212-017-0341-1

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*, 506-519. doi:10.1016/j.learninstruc.2010.08.002

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. doi:10.1080/03055698.2012.654920

Wooley, R., Was, C., Schunn, C. D., & Dalton, D. (2008, July). *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. doi:10.1016/j.iheduc.2008.06.005

7

English summary

INTRODUCTION

Giving feedback to peers can be a powerful learning tool. When students give feedback on peers' learning products, they become cognitively involved with the material, as they need to think about the key features of the product and suggest ways to improve it. This type of thinking can lead to learning by feedback providers (who are referred to as reviewers in this thesis).

In trying to maximize the learning potential of the feedback-giving process, it is necessary to study that process as well as the factors that can influence it. According to the literature, giving feedback to peers includes three steps: 1) defining assessment criteria, 2) judging peers' products or performance, and 3) providing feedback for future learning. Each step can be organized in different ways, which can influence the learning that originates from giving feedback. First, when giving feedback, students can either use assessment criteria given by a teacher or think of their own criteria. The cognitive processes involved can differ, which may lead to different learning outcomes. Second, the quality and type of products reviewed can influence the feedback given by students, and different feedback may require different types of thinking, which, in turn, may lead to different learning. Finally, the way feedback is given, for example, by commenting or by grading, might also influence any learning gain, as students could be stimulated to invest different level of mental effort in different feedback methods.

This thesis aimed at investigating factors that can influence reviewers' learning by conducting experimental studies, with each study focusing on one of the steps of the feedback-giving process.

OVERVIEW OF THE STUDIES THAT WERE CONDUCTED

All studies were conducted with secondary school students (14-15 years old) from either Russia or the Netherlands. During the experimental lessons, students worked in online inquiry learning environments that were created using the Go-Lab ecosystem (www.golabz.eu), which are called inquiry learning spaces (ILSs). The topics of the experimental lessons were related to the national science curricula of the two countries.

The outline of all experiments was the same: pre-testing of domain knowledge, followed by a lesson with an ILS that included creating students' own concept map and giving feedback on (fictitious) peers' products, and, finally, post-testing of domain knowledge. The knowledge tests consisted of open-ended questions and were graded by the researcher and a second rater using a coding scheme that was developed. The quality of students' own concept maps and the quality of the feedback they provided were also coded by two raters and included in the analysis.

Giving feedback was done anonymously in a specially designed online tool (the Peer Assessment Tool). The configuration of the tool differed per study depending on the research questions. The tool always presented the products to review, and guided students through the process of giving feedback either by presenting pre-determined assessment criteria (Studies 1, 3, and 4) or by encouraging students to think of their own criteria (Study 2, one condition). Assessment criteria in all studies matched the type of product reviewed. The products to give feedback on were relatively small-scale compared to reports, essays or papers, which are often used in this line of research. This also meant that the feedback-giving moment could be rather brief. Using these smaller scale products means that the results found are easier to transfer to practical teaching situations.

Study 1

The first study focused on the third step of the feedback-giving process and investigated the role that the feedback method plays in reviewers' learning, by comparing commenting and grading. Previous research had indicated that providing comments on peers' work could be more beneficial for reviewers' learning than grading. This could be because commenting requires more cognitive investment by the reviewer than grading. However, it was not clear if the same trend would be found when the products reviewed are relatively small, as compared to essays or papers.

To investigate the influence of the feedback method on learning, the participants in this study, coming from Russia and the Netherlands, gave feedback on two concept maps by either writing comments or grading the concept maps with smileys. The results showed that students from the commenting condition had higher post-test scores (taking students' prior knowledge into account) than students from the grading condition. Moreover, an interaction effect was found for students' prior knowledge and the feedback method, such that students with low prior knowledge benefited the most from commenting. The quality of students' own concept maps predicted their post-test scores, with a higher number of correct links predicting higher post-test scores.

Study 2

The second study investigated the effect that being provided with assessment criteria had on reviewers' learning, which means that the study concerned the first step in the process of giving feedback. The literature presents inconclusive results about the role of provided assessment criteria. Some research has argued that assessment criteria can guide students through the potentially challenging process of giving feedback, thus implying a positive influence of provided assessment criteria on reviewers' learning. Other research has suggested that using their own assessment criteria might be more useful for students' learning, as they can fully understand and apply these criteria, which might not be the case for the provided criteria.

To study the effect of providing assessment criteria, participants in this study (Russian secondary school students) gave feedback on two concept maps. They were assigned to one of two conditions: they either were given assessment criteria or had to think of their own assessment criteria. The results did not reveal any difference in post-test scores (taking students' prior knowledge into account) between conditions. Students who were not provided with assessment criteria were still able to give meaningful content-related feedback. Moreover, for both conditions the quality of the feedback given by students predicted their post-test score, with better feedback predicting higher scores.

Study 3

The third study was connected with the second step in the feedback-giving process: judging peers' products. The aim was to investigate whether the quality of the product that was reviewed influenced reviewers' learning. According to the literature, products with different levels of quality require different feedback, which could lead to different learning for the reviewers. It was not clear which would trigger more learning: studying good examples in higher quality products or finding mistakes in lower quality products; the situation was especially unclear for smaller scale products.

To investigate the role of the quality of the reviewed product, Dutch secondary school students were assigned to give feedback on a pair of concept maps with particular levels of quality, which differed in the number of included concepts and mistakes. There were three sets used: two concept maps of lower quality (condition 1), one concept map of lower quality and one concept map of higher quality (condition 2), two concept maps of higher quality (condition 3). The results showed that students who gave feedback on lower quality concept maps had higher post-test scores (taking students' prior knowledge into account) than students who gave feedback on higher quality concept maps. For all conditions, students who changed their own concept map after giving feedback had higher post-test scores than students who did not make changes.

Study 4

Like the previous study, the fourth study concerned the product that was reviewed, but instead of different levels of quality, it compared the learning resulting from reviewing two different types of product: concept maps and answers to open-ended test questions. The literature suggested that, similarly to the product's level of quality, the type of product reviewed could influence reviewers' learning. On the one hand, concept maps could trigger more conceptual thinking than answers to test questions, which would be beneficial for learning. On the other hand, concept maps might also be too challenging to review for particular groups of students, which would prevent them from learning.

To study the effect of the type of product on reviewers' learning, Dutch secondary school students were asked to give feedback on either two concept maps or a set of answers to open-ended test questions. Both types of product were comparable in terms of the content: they included the same misconceptions and were of average quality. The results did not reveal a difference in post-test scores (taking students' prior knowledge into account) between the conditions. Higher quality feedback was provided on answers to test questions than for concept maps, and for both conditions, the quality of feedback given by students predicted their post-test scores.

Conclusion

Traditionally, peer feedback has been used as a replacement for feedback from teachers; however, it is more and more being viewed as a learning tool rather than an assessment tool. Knowing more about the factors that can influence reviewers' learning can contribute to better understanding of the feedback-giving process and how to organize it to enhance learning.

Based on the results of the studies, several conclusions can be drawn. First, giving feedback by commenting seemed more beneficial for reviewers' learning than giving feedback by grading, especially for lower prior knowledge students. Second, secondary school students could give meaningful content-related feedback even without being provided with assessment criteria. Third, the quality and type of reviewed products could influence reviewers' learning. Product features that could stimulate the most learning from reviewing smaller scale products include familiarity of the product and a level of quality that is the same as or lower than the reviewer's current level of performance. Finally, the quality of feedback provided by students can be an important indicator of their learning, so students should be encouraged to invest in giving high-quality feedback. Investigating how to encourage students to provide good feedback could be an interesting direction for future research. Another direction could be to see if the results generalize to other types of smaller scale products (e.g., hypotheses).

Generally, the studies that were conducted showed that with proper organization, even a brief moment of feedback on a small-scale product could be a learning experience for a feedback provider; therefore, giving feedback on peers' work could be used in secondary (science) education more often.

8

Nederlandse samenvatting

INLEIDING

Het geven van feedback aan *peers* (klasgenoten) kan een krachtig leermiddel zijn. Wanneer leerlingen feedback geven op leerproducten van *peers* zijn ze cognitief bezig met het materiaal, omdat ze moeten nadenken over de belangrijkste kenmerken van het product en manieren moeten voorstellen om het product te verbeteren. Dit denkproces kan leiden tot leren bij degene die feedback geeft (die in dit proefschrift een *reviewer* wordt genoemd).

Om het leren bij het proces van feedback geven te maximaliseren, is het van belang om zowel dit proces als factoren die mogelijk van invloed kunnen zijn te onderzoeken. Volgens de literatuur bestaat het geven van feedback aan *peers* uit drie stappen: 1) definiëren van beoordelingscriteria, 2) beoordelen van producten of prestaties van *peers* en 3) het geven van feedback voor toekomstig leren. Elke stap kan op verschillende manieren worden vormgegeven, hetgeen van invloed kan zijn op het leren dat voortkomt uit het geven van feedback. Ten eerste kunnen leerlingen, wanneer ze feedback geven aan hun *peers,* gebruik maken van beoordelingscriteria die door een docent zijn opgesteld of hiervoor hun eigen criteria bedenken. Deze twee cognitieve processen kunnen verschillen, wat kan resulteren en verschillende leeruitkomsten. Ten tweede kan de kwaliteit en het type beoordeelde product van invloed zijn op de feedback die leerlingen geven. Verschillende soorten feedback kunnen verschillende manieren van denken vereisen, hetgeen vervolgens kan leiden tot een verschil in leren. Ten slotte kan de manier waarop feedback wordt gegeven, bijvoorbeeld het plaatsen van opmerkingen of het geven van een score, ook van invloed zijn op leeruitkomsten, aangezien leerlingen bij verschillende manieren van feedback geven mogelijk tot een verschillende mate van mentale inspanning aangezet worden.

Dit proefschrift is op gericht op het onderzoeken van factoren die het leren van de *reviewer* kunnen beïnvloeden. Dit is gedaan door middel van experimentele studies waarbij elke studie zich focust op een van de stappen van het proces van het feedback geven.

OVERZICHT VAN DE UITGEVOERDE STUDIES

Alle studies zijn uitgevoerd met middelbare scholieren (14-15 jaar oud), afkomstig uit Rusland of Nederland. Tijdens de lessen waarin het experiment plaatsvond, hebben leerlingen gewerkt in online onderzoekende leeromgevingen, genaamd ILS'en (*Inquiry Learning Spaces* in het Engels), die zijn ontworpen met het *Go-Lab ecosystem* (www.golabz.eu). De onderwerpen die tijdens de lessen van het experiment aan bod kwamen, waren in beide landen gerelateerd aan hun curricula van de exacte vakken.

De opzet van alle experimenten was hetzelfde: het vooraf toetsen van de domeinkennis, gevolgd een les, waarin leerlingen in een ILS onder andere een concept map maakten en feedback gaven op producten van (fictieve) *peers*, en ten slotte het achteraf toetsen van de domeinkennis. Kennistoetsen bestonden uit open vragen en werden door de onderzoeker en een tweede beoordelaar beoordeeld aan de hand van een daarvoor ontwikkeld codeerschema. De kwaliteit van de concept map die de leerlingen zelf gemaakt hadden en de kwaliteit van de feedback die ze hebben gegeven werd ook gecodeerd door twee beoordelaars en meegenomen in de analyse.

Het geven van feedback gebeurde anoniem in een daar speciaal voor ontworpen online tool (*Peer Assessment Tool* in het Engels). De configuratie van de tool was afhankelijk van de onderzoeksvragen en verschilde per studie. De tool toonde altijd de te beoordelen producten en begeleidde de leerlingen in het proces van feedback geven door ofwel beoordelingscriteria weer te geven (Studie 1, 3 en 4), of door leerlingen te stimuleren om zelf criteria te bedenken (Studie 2, één conditie). In alle studies zijn de beoordelingscriteria aangepast aan de beoordeelde producten. De producten waar feedback op werd gegeven waren kleiner van omvang in vergelijking met bijvoorbeeld verslagen, essays of papers, waardoor ook het moment van feedback geven relatief kort kon zijn. Vanwege het gebruik van deze kleinschaligere producten kunnen de gevonden resultaten makkelijker vertaald worden naar praktische onderwijs situaties.

Studie 1

In de eerste studie lag de focus op de derde stap van het proces van feedback geven en werd de rol die de manier van feedback geven speelt op het leren van de *reviewer* onderzocht. Meer specifiek is hier het plaatsen van opmerkingen vergeleken met het geven van een score. Resultaten uit eerder onderzoek suggereren dat het plaatsen van opmerkingen bij het werk van *peers* bevorderlijker kan zijn voor het leren van de *reviewer* dan het geven van een score. Dit kan worden verklaard door het feit dat het plaatsen van opmerkingen mogelijk meer cognitieve inspanning van de *reviewer* vereist dan het geven van een score. Het was echter nog niet duidelijk of eenzelfde trend zichtbaar zou zijn in een situatie waarin de beoordeelde producten relatief klein zijn in vergelijking met essays of papers.

Om de invloed van de manier van feedback geven op het leren te onderzoeken gaven participanten van deze studie, afkomstig van Rusland en Nederland, feedback op twee concept maps door ofwel geschreven opmerkingen te plaatsen of door het geven van een score in de vorm van smileys. De toewijzing van leerlingen aan een van deze twee condities gebeurde willekeurig. De resultaten toonden aan dat de leerlingen uit de opmerkingen conditie hogere natoetsscores hadden dan leerlingen uit de score conditie (rekening houdend met de voorkennis van de studenten). Daarnaast werd een interactie-effect gevonden van voorkennis van leerlingen en de manier van feedback geven, waarbij leerlingen met een lage voorkennis het meeste profiteerden van het plaatsen van opmerkingen. De kwaliteit van de concept maps die door de leerlingen zelf gemaakt was voorspelde hun natoetsscores, waarbij een hoger aantal correcte connecties leidde tot hogere natoetsscores.

Studie 2

In de tweede studie werd het effect van het voorzien worden van beoordelingscriteria op het leren van de *reviewers* onderzocht, wat betekent dat de studie betrekking had op de eerste stap in het proces van het geven van feedback. De literatuur laat geen eenduidige resultaten zien als het gaat om de rol van vooraf vastgestelde beoordelingscriteria. In sommige onderzoeken werd gesteld dat beoordelingscriteria begeleiding kunnen bieden tijdens het mogelijk uitdagende proces van feedback geven, wat aantoont dat het geven van vooraf vastgestelde beoordelingscriteria een positieve invloed kan hebben op het leren van de *reviewer*. Ander onderzoek suggereerde dat het gebruik van eigen beoordelingscriteria nuttiger zou kunnen zijn voor het leren van leerlingen, aangezien zij deze criteria volledig kunnen begrijpen en toepassen, wat wellicht niet het geval is bij vooraf vastgestelde criteria.

Om het effect van vooraf vastgestelde beoordelingscriteria te onderzoeken gaven participanten van deze studie (Russische middelbare scholieren) feedback op twee concept maps. Ze werden daarvoor toegewezen aan een van de twee condities: ze konden ofwel vooraf vastgestelde beoordelingscriteria gebruiken of ze moesten hun eigen beoordelingscriteria bedenken. De resultaten lieten geen verschil zien in natoetsscores tussen condities (rekening houdend met de voorkennis van de studenten). Leerlingen die geen vooraf vastgestelde beoordelingscriteria ontvingen, waren nog steeds in staat om zinvolle inhoudelijke feedback te geven. Daarnaast voorspelde de kwaliteit van de feedback die leerlingen gaven in beide condities hun natoetsscore, waarbij betere feedback leidde tot hogere scores.

Studie 3

De derde studie hield verband met de tweede stap van het proces van feedback geven: het beoordelen van producten van *peers*. Het doel was om te onderzoeken of de kwaliteit van het beoordeelde product invloed had op het leren van de *reviewers*. Op basis van de literatuur zouden beoordeelde producten van verschillende kwaliteit verschillende feedback vereisen, wat zou kunnen leiden tot verschil in het leren van de *reviewer*. Het was daarbij niet duidelijk wat meer leren teweeg zou kunnen brengen: het bestuderen van goede voorbeelden in producten van hogere kwaliteit of het vinden van fouten in producten van lagere kwaliteit; dit bleek vooral onduidelijk voor kleinschaligere producten.

Om de rol van de kwaliteit van het beoordeelde product te onderzoeken, werd aan Nederlandse middelbare scholieren een set concept maps van een bepaalde kwaliteit toegewezen om die te beoordelen. Deze concept maps verschilden voor wat betreft het aantal concept en fouten. Er werd gebruik gemaakt van drie sets: twee concept maps van lagere kwaliteit (conditie 1), één concept map van lagere kwaliteit en één concept map van hogere kwaliteit (conditie 2) en twee concept maps van hogere kwaliteit (conditie 3). De resultaten toonden aan dat leerlingen die feedback gaven op concept maps van lagere kwaliteit hogere natoetsscores hadden dan leerlingen die feedback gaven op concept maps van hogere kwaliteit (rekening houdend met de voorkennis van de studenten). Voor alle condities gold dat leerlingen die hun eigen concept map hadden veranderd nadat ze feedback hadden gegeven, hogere natoetsscores hadden dan leerlingen die hem niet hadden veranderd.

Studie 4

Net als de vorige studie had de vierde studie betrekking op het beoordeelde product, maar in plaats van het vergelijk van verschillende kwaliteitsniveaus werd in deze studie het leren als gevolg van het beoordelen van twee type producten vergeleken: concept maps en antwoorden open toetsvragen. In de literatuur wordt gesuggereerd dat, net als het kwaliteitsniveau van het product, het type beoordeelde product van invloed kan zijn op het leren van de *reviewer*. Aan de ene kant zouden concept maps een hogere mate van conceptueel denken kunnen uitlokken dan antwoorden op toetsvragen, hetgeen gunstig kan zijn voor het leren. Aan de andere kant kunnen voor bepaalde groepen leerlingen concept maps mogelijk een te grote uitdaging vormen om te beoordelen, hetgeen hun leren zou kunnen belemmeren.

Om het effect van het type product op het leren de *reviewer* te onderzoeken, werd aan Nederlandse middelbare scholieren gevraagd om feedback te geven op ofwel twee concept maps of op antwoorden op open toetsvragen. Beide typen producten waren qua inhoud vergelijkbaar: ze bevatten dezelfde misvattingen en waren van gemiddelde kwaliteit. De resultaten lieten geen verschil in natoetsscores tussen de condities zien (rekening houdend met de voorkennis van de studenten). De feedback die gegeven werd op antwoorden van toetsvragen was van hogere kwaliteit dan de feedback op concept maps. In beide condities voorspelde de kwaliteit van de feedback die leerlingen gaven hun natoetsscores.

CONCLUSIE

Oorspronkelijk werd *peer*feedback gebruikt ter vervanging van feedback gegeven door leerkrachten, maar het wordt inmiddels steeds vaker gezien als een leermiddel in plaats van een beoordelingsmiddel. Meer kennis over de factoren die het leren van de *reviewer* kunnen beïnvloeden, kan helpen om het proces van het geven van feedback beter te begrijpen en om het zo te organiseren dat het bevorderlijk is voor het leren.

Op basis van de resultaten van de studies kunnen verschillende conclusies worden getrokken. Ten eerste leek het geven van feedback door middel van het plaatsen van opmerkingen bevorderlijker voor het leren van de *reviewer* dan het geven van feedback door middel van het geven van een score, vooral voor leerlingen met minder voorkennis. Ten tweede konden leerlingen uit het voortgezet onderwijs zinvolle inhoudelijke feedback geven, zelfs zonder dat ze daarbij vooraf vastgestelde beoordelingscriteria ontvingen. Ten derde kunnen de kwaliteit en het type van de beoordeelde producten van invloed zijn op het leren van de *reviewer*. Kleinschaligere producten die het leren het meest kunnen stimuleren leken met name producten te zijn waar de *reviewer* bekend mee is en die van eenzelfde of lager niveau zijn dat wat een *reviewer* zelf zou kunnen. Ten slotte kan de kwaliteit van de feedback die door leerlingen wordt gegeven een belangrijke indicatie zijn voor hun leren, wat betekent dat leerlingen gestimuleerd zouden moeten worden om feedback van hoge kwaliteit te geven. Onderzoek naar hoe leerlingen kunnen worden gestimuleerd om goede feedback te geven kan een interessante richting zijn voor toekomstig onderzoek. Een andere richting kan zijn om te onderzoeken of de verkregen resultaten ook gelden voor andere soorten kleinschalige producten (bijvoorbeeld hypotheses).

Over het algemeen laten de uitgevoerde studies zien dat, mits goed georganiseerd, zelfs een kort feedbackmoment op een kleinschalig product een leerzame ervaring zijn voor degene die feedback geeft. Om die reden zou het geven van feedback op producten van *peers* vaker ingezet kunnen worden in het voortgezet onderwijs.

Giving feedback can be a powerful learning tool when properly organized.
This dissertation seeks ways to maximize students' learning from giving feedback.

Four experimental studies investigated how students' learning is influenced by the way they give feedback, the origin of assessment criteria, and the quality level and type of the reviewed products.

Results indicated that even a brief moment of giving feedback can be beneficial for students' learning and they can guide practitioners in designing the process of giving feedback.