JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Ketonen, Laura; Hähkiöniemi, Markus; Nieminen, Pasi; Viiri, Jouni

Title: Pathways Through Peer Assessment : Implementing Peer Assessment in a Lower Secondary Physics Classroom

Year: 2020

Version: Published version

Copyright: © The Author(s) 2020

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Ketonen, L., Hähkiöniemi, M., Nieminen, P., & Viiri, J. (2020). Pathways Through Peer Assessment : Implementing Peer Assessment in a Lower Secondary Physics Classroom. International Journal of Science and Mathematics Education, 18(8), 1465-1484. https://doi.org/10.1007/s10763-019-10030-3

International Journal of Science and Mathematics Education
https://doi.org/10.1007/s10763-019-10030-3

Pathways Through Peer Assessment: Implementing Peer Assessment in a Lower Secondary Physics Classroom

Laura Ketonen[1] · Markus Hähkiöniemi[1] · Pasi Nieminen[1] · Jouni Viiri[1]

Received: 4 June 2019 / Accepted: 16 October 2019/Published online: 02 January 2020
© The Author(s) 2020

Abstract
Peer assessment has been shown to advance learning, for example, by improving one's work, but the variance of learning benefits within or between studies has not been explained. The purpose of this case study was to examine what kinds of pathways students have through peer assessment and to study which factors affect them when peer assessment is implemented in the early stage of physics studies in the context of conducting and reporting inquiry. Data sources used include field notes, audio recordings of lessons, student lab reports, written peer feedback, and student interviews. We examined peer assessment from the perspective of individual students and found 3 profiles of peer assessment: (1) students that improved their lab report after peer assessment and expressed other benefits, (2) students that did not improve their lab report but expressed other benefits, and (3) students that did not experience any benefits. Three factors were found to explain these differences in students' pathways: (1) students' engagement in conducting and reporting inquiry, (2) the quality of received feedback, and (3) students' understanding of formative assessment. Most students experienced some benefits of peer assessment, even if they did not put effort into their own work or receive constructive feedback. Nevertheless, in this case study, both improving one's work and experiencing other benefits of peer assessment required sufficient accomplishment of all 3 factors.

Keywords Case study · Formative assessment · Peer assessment · Physics learning · Secondary school

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10763-019-10030-3) contains supplementary material, which is available to authorized users.

✉ Laura Ketonen
laura.k.ketonen@jyu.fi

[1] Department of Teacher Education, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland

Introduction

Formative assessment (FA) has attracted significant interest among educators and researchers since Black and Wiliam published their seminal articles in (1998a, b). While the reported results have been criticized (Bennett, 2011; Kingston & Nash, 2011), the benefits of FA are not in question. Black and Wiliam (1998b) define assessment as formative when the information gathered in the assessment is "actually used to adapt the teaching to meet student needs" (p. 140). This definition does not specify who adapts the teaching, but we deduce that it is not only the teacher but also students and their peers. By definition, an adaptive process must consider the circumstances. Consequently, the research of FA should encompass knowledge of individual students in order to describe its very essence. In this study, we will investigate peer assessment from this angle.

According to the definition of Black and Wiliam (1998b), practically all classroom activities can be used for FA, including classroom discussion, observation, tests, projects, and self- and peer assessment. Peer assessment (PA) can be defined as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners" (Topping, 2009, p. 20) and is used for formative or summative purposes (Topping, 2013). In this paper, a formative view is taken, in which students help their peers to move along the learning pathway by bridging the gap between their actual level of achievement and their goals (Sadler, 1989).

The research of PA in science education has not examined what contributes to a student's pathway at the level of the individual. We will take up the challenge and analyze students' pathways by carefully considering where students begin with the original work, as well as how they give and receive feedback and the benefits that seem to be part of the PA process. Despite the gap in research, there is knowledge of outcomes of PA and the factors influencing them; next, we will introduce them. We refer mostly to studies of science education in secondary schools, which is the framework of this study, but since PA is under-researched at the secondary level (van Zundert, Sluijsmans, & van Merriënboer, 2010; Topping, 2013), we also include a few other studies outside this definition. It is not possible to strictly define which phenomena of peer assessment are specific to a school subject or to what degree. Gotwals, Philhower, Cisterna and Bennett (2015, p. 421) discuss the same dilemma relating generally to FA saying that their "findings suggest there are some aspects of expertise in FA that are applicable and can be analysed across disciplines, but that it is important to examine teachers' practices in different disciplines to fully characterize expertise in FA." We see that this describes the case of PA as well.

Outcomes and Contributing Factors of Peer Assessment

Tsivitanidou, Zacharia, and Hovardas (2011) concluded that seventh-grade students with no training or experience with PA had beginning skills of PA and were able to produce peer feedback, but the validity and reliability of the feedback in this study were low. The low level of peer feedback comments of seventh-grade students was also noticed by Tasker and Herrenkohl (2016). According to them, guiding students to reflect on the elements of meaningful feedback by using the peer feedback comments of previous PA helped them to become conscious of the qualities of constructive and

not constructive feedback. This resulted in more significant feedback during the next round.

Students' good skills in the subject in question seem to facilitate the provision of good-quality feedback. Chong (2017) found that seventh-grade students' own writing abilities affected the relevancy and accuracy of the feedback they produced. Lu and Law (2012) found that for 13- and 14-year-old high school students on a liberal studies course, peer feedback that identified problems and gave suggestions predicted good performance of the assessor. On the other hand, positive affective feedback was related to the assessee's good performance, which means that better quality work induced more positive affective comments.

The study of Anker-Hansen and Andrée (2019) shows that providing feedback can be equally effective as receiving it. They researched eighth- and ninth-grade science students making an inquiry, and found that similar amounts of improvements in students' work were due to received and given feedback. Liu and Carless (2006) claim that one of the advantages of PA is engaging students with criteria representing the standards. Chetcuti and Cutajar (2014) concluded in their research of PA in post-secondary physics that students experienced PA as a learning tool that helped them better understand the learning outcomes and criteria of high-quality work. Understanding criteria may contribute success to either the work at hand, subsequent tasks, or both. The quality of received feedback is related to the benefits of PA. Tseng and Tsai (2007) found that 10th graders significantly improved their computer course projects after three rounds of PA. They reported that different types of feedback had different effects on students' subsequent learning. Reinforcing feedback was most useful for developing students' subsequent projects. Suggestive feedback was helpful on the first round of PA but not on later rounds, whereas corrective feedback with a lecture tone was noticed to have a negative effect on subsequent work. Gielen, Peeters, Dochy, Onghena, and Struyven (2010) noticed that "justified comments" improved seventh-grade students' writing performance, but the effect diminished with more skilled students. Justification was noticed to be more important than the accuracy of feedback. These studies show that the quality of given and received peer feedback is an important factor for benefits, but again more knowledge is needed at the individual level.

Formative PA is often used to help students improve their work. Starting with more modest results, Tsivitanidou, Zacharia, Hovardas, and Nicolaou (2012) researched eighth graders who received peer feedback using a computer-supported inquiry learning environment and found that none of them made subsequent changes to their science-related work. On a seventh-grade science project, Tsivitanidou et al. (2011) reported that a minority of students made changes to their portfolios after receiving peer feedback, though they do not specify whether the changes were actual improvements. More positively, Anker-Hansen and Andrée (2019) reported four out of five eighth- and ninth-grade science students revising their work after giving and receiving peer feedback and discussing it in small groups. Tsivitanidou, Constantinou, Labudde, Rönnebeck, and Ropohl (2018) researched upper secondary school students using PA in physics and reported 10 out of 11 pairs revised their models after PA; of those 10, all improved the quality of their initial models and most of them rose to a higher level of attainment.

As the above studies show, PA has the potential to help students improve their work, but more knowledge is needed to explain the variation in these results. Why do some students benefit more from PA than others?

This Study

A more holistic, qualitative view is needed to understand the mechanisms of peer feedback. In order to successfully implement PA, we need knowledge of the pathways of individual students. Who does and does not benefit from PA, and why? Students may lack skills in many areas, but which are the real bottlenecks? In this study, we examine students' pathways through PA and describe the factors that facilitate or hinder the benefits. The research questions are as follows:

1. What kind of pathways do students take when peer assessment is implemented in a lower secondary school physics classroom?
2. Which factors advance or reduce the benefits of peer assessment?

Method

Participants

This study was carried out at a lower secondary school in Finland. The participants were two classes of seventh-grade students (n = 29; mean age 13 years) taught by the same teacher. The permission for the research was granted by the educational administration of the city. Students' parents signed a consent form that allowed us to gather the data. The actual PA required attendance in three subsequent lessons; seven students missed one or more, which decreased the number of participants to 22.

Finland has a strong tradition of summative assessment. In the National Core Curriculum (NCC), the concept of FA was first introduced as an addition in 1999 (Finnish National Board of Education (FNBE), 1999), and the recent NCC (FNBE, 2014) stated for the first time that the emphasis of assessment should be on FA. In the same NCC, PA was mentioned for the first time. Students in this study changed to the new NCC in fifth grade. As teaching practices follow curricula changes slowly, students had little experience with PA.

Procedure

This was students' first course in physics at a lower secondary school. It included the basics of physics inquiry, mechanics, and dimensions of the universe. In primary school, students had studied general science, which includes biology, geography, physics, chemistry, and health education, and is taught by a class teacher. Their studies do not include a quantitative approach to physics, like calculating speed. Few students had any memory of inquiry activities, and even those activities appeared as cookbook experiments.

During their physics course, students were trained in PA: understanding the elements of good feedback and its purpose, understanding criteria-based assessment, and comparing a work to criteria (Supplementary material). The researcher, who was an experienced teacher with inside knowledge of the school, planned the training, which included class discussions, written tasks, and actual peer assessments between groups and individuals. There were six 10- to 45-min training sessions over 6 weeks' time before the PA described in this paper. The researcher organized the training and PA in cooperation with the teacher and participated in the majority of lessons as an observer and assistant teacher.

Earlier in this course, students had built a "Mars rover" as a technology project. The instruction had been to make a vehicle that moves on its own. Students could use any available material to create the movement, such as rubber bands, balloons, or simple electric motors with batteries. The task, which later was peer assessed, was to produce a lab report determining the speed of the rover. The inquiry was conducted in groups of three to four students, but each student produced their own report. Two lessons (1.5 h each) were dedicated to conducting the inquiry, but as anticipated, adjusting and fixing the rovers took some of the time. The reports were to be graded by the teacher, accounting for a sixth of the course marks.

Before returning their reports, students were invited to assess the quality of another student's product according to the criteria provided and to revise their own work afterwards. Using a report as a basis of assessment is supported by Emden and Sumfleth (2016), who claim that assessing a report captures "a broad continuum of student's achievements." Students assessed their peer's work using a three-choice rubric with assessment criteria and an opportunity to include written comments (Fig. 1). The criteria resembled the process of a scientific inquiry, which was provided to students before the inquiry task and explained by the teacher. Before PA, the teacher read and explained the criteria and the use of the assessment scale. She also helped students by discussing the criteria with individual students during the assessment.

The researcher planned the pairing of students with the teacher. Students in the same inquiry group would not assess each other's work, but social factors were also considered. For example, the work of a timid student was not given to a loud bully. Students pairs were "equal-status" in a broad meaning of the word, meaning that their role in the classroom was the same, but not necessarily that their cognitive or social skills were equal. Students had 45 min to produce peer feedback and after that, another 45 min to read the received feedback and rework their own lab report. Here, the formative nature of peer feedback emerged; it was used to improve one's work and understanding. This was stressed to students: feedback, confirming or corrective, was given in order to help other students. Students were advised to receive the feedback with the same mindset—to use what was helpful and ignore what was not. After rework, lab reports were returned to the teacher for summative assessment.

Research Design

Since we wanted to explain the outcomes of PA, a case study design was adopted. It served our goals and research questions well. Jindal-Snape and Topping (2010) state that "The purpose of a case study is to get in-depth information regarding what is

criteria	meeting the criteria			written comment
Research plan and equipment have been clearly presented.	☹	😐	☺	
Measurements are reasonable and clearly presented.	☹	😐	☺	
Report shows, how speed is calculated from measurements. The calculation includes an equation, numbers and units.	☹	😐	☺	
Error of measurements and results have been commented.	☹	😐	☺	

Figure 1 Model of peer assessment sheet

happening, why it is happening and what the effects are of what is happening" (p. 20). We wanted to complement the research of PA by taking a qualitative approach to it.

Data Collection

The data collected during the 10-week course (20 sessions of approximately 90 min each) included the researcher's field notes, audio recordings of lessons, students' original lab reports, written peer feedback, revised lab reports, and interviews.

The researcher made field notes of all the lessons she participated in. Field notes were "issue oriented" (Hopkins, 2008, p. 105), concentrating on observations and reflections of PA. Observations of producing and assessing lab reports can be considered focused (Hopkins, 2008, pp. 88–89) since besides general observations, a five-scale pre-made rubric was used to describe students' engagement with the task.

All 22 students were individually interviewed in the same week they returned their lab reports. In semi-structured, stimulated recall interviews (Ryan & Gass, 2012), students were shown copies of their original and revised lab reports and the feedback they received. The interview focused on students' thoughts about the original work, the received feedback, and their reasoning for making changes (or not). Students' views of the benefits of PA were also inquired. To introduce the issue of providing feedback, each student was shown a copy of the peer's lab report and the feedback they provided.

Analysis

To answer the first research question, students' pathways through PA were analyzed at five stages, which are presented in Table 1. Next, the analysis is explained in detail.

Students' original lab reports were assessed by an expert, whose assessment was compared to the peer's assessment. The expert was a researcher with several years of teaching experience at a secondary school. Each criterion (Fig. 1) was coded regarding

Table 1 Stages of peer assessment and data sources

Stage	Explanation	Data sources
1. Original work	Did student put effort into original work (conducting and reporting inquiry)?	Field notes, with interviews and audio files for triangulation
2. Providing feedback	What was the quality of feedback student provided of peer's work?	Original lab reports, feedback sheets
3. Receiving feedback	What was the quality of received peer feedback and what were student's thoughts of it?	Original lab reports, feedback sheets
4. Improving work	Did student improve his/her work after PA?	Original and revised lab reports
5. Other benefits	Did student experience other benefits of PA, besides possible changes?	Interviews

three aspects (Table 2): (1) Did the assessee fulfill a single assessment criterion, (2) did the assessor notice whether the criterion was fulfilled, and (3) had the assessor supplied a written comment and, if they had, was it constructive or not? Comments like "well done" or "yes" were categorized as unconstructive since they did not add anything to the mark. A constructive comment about a correct answer described in detail what was good, such as "Results are well explained." We call this *a constructive compliment*. A constructive comment about an incorrect answer clarified what was incorrect or gave guidance, such as "The stopwatch has not been mentioned." We call this *a constructive critique*.

Two other researchers assessed five students' lab reports to test the reliability of the expert's assessment. The agreement between these two assessors and the expert was 62.5% for the first two students (8 criteria), but after discussion and the expert explaining her interpretation of the criteria, the agreement of the last three students (12 criteria) was 92%. The constructiveness of feedback was tested with the same two coders with an agreement of 80% (agreement in 16 of 20 comments). The differences in coding were then discussed and agreed upon. With increased understanding, the expert recoded the remaining 17 cases, but ended up making no changes.

After coding each criterion, we looked at the quality of the feedback (all four criteria, Fig. 1) from the assessor's point of view. From that angle, it is not reasonable to separate constructive feedback about something that could be improved from something that was well done since the quality of assessed work affects the quality of feedback. If the work is perfectly done, no constructive critique can be given, and if the work is very inadequate, constructive compliments are not appropriate. Hence, we divided the provided feedback into only two categories. If the student had provided one or more constructive comments, the feedback was considered constructive since it had the capacity to help the assessee to recognize either his or her strengths or development points. If there were no constructive comments, feedback was considered not constructive. The codes are presented in Table 3.

The received feedback of a single student was categorized into four groups (Table 3): Feedback that included (1) both constructive critique and constructive compliments, (2) constructive critique, (3) constructive compliments, or (4) no constructive comments.

The changes in students' reports were also marked. The same researcher investigated whether they had improved the quality of their lab report in terms of the given criteria. Cosmetic changes and no changes at all were considered as a single group of "no improvement." If a student raised the level of his/her work in terms of the criteria, it was categorized as an "improvement" (Table 3). Again, two other researchers assessed

Table 2 Coding the quality of assessment of single criteria

Did the assessee fulfill the criterion?	Did the assessor notice whether the criterion was fulfilled?	Is there a written comment besides marking?
Yes	Yes	No
No	No	Yes, an unconstructive comment
		Yes, a constructive comment

Table 3 Alternatives of coding a student's performance on each stage of PA

Sequence	Values	Definition
Original work	Effort	No signal of lack of effort
	Lack of effort	Distinct signal of lack of effort
Providing feedback (assessor's angle)	Constructive	Constructive comments (compliment or critique)
	Not constructive	No constructive comments
Receiving feedback (assessee's angle)	Constructive critique and compliment	Both constructive critique and constructive compliment
	Only constructive critique	One or more constructive critique but no constructive compliment
	Only constructive compliments	One or more constructive compliment but no constructive critique
	Not constructive	No constructive comments
Improving work	Improvement	Student has improved the quality of his/her work in terms of assessment criteria
	No improvement	Student has not improved the quality of his/her work in terms of assessment criteria
Other benefits	Other benefits	Student brings up other personal benefits of PA besides improving his/her work
	No other benefits	Student does not bring any other personal benefits of PA besides improving his/her work

the original and revised lab reports of five students who had made changes. There was an 80% agreement (four of five cases).

Students' interviews were transcribed and sections regarding benefits of PA were open coded. Codes were used to retrieve and organize the chunks of data (Miles & Huberman, 1994) and from them categories were formed. Categories were named and their properties specified (Strauss & Corbin, 1998). In the end, interviews were read through and categorization of every piece was rechecked. Four categories of benefits were found besides improving work: (1) reinforcing one's own capability, (2) confirming that some part of one's own work is well done, (3) learning something for the future, and (4) improving one's mood. With two other assessors, an agreement of benefits was found in four of five cases (80%). Both assessors thoroughly reflected on the disagreed work, and in discussion afterwards, all three assessors agreed that it was a borderline case. The categories were noticed to be appropriate.

Field notes and interviews were open coded regarding effort with the original work. Focus was on the researcher's five-scale observations of students working. Research shows that teachers' ratings of students' efforts correlate positively with student reports (Zhu & Urhahne, 2014), but in order to avoid misinterpretations, we considered only the students with the lowest observer marks as having a lack of effort. The description of the level was "Student tries to avoid the task. Work proceeds only when motivated/pushed by teacher or not even then."

Students' PA experiences (Table 3) were represented as pathways from the individual student's point of view. In order to answer the first research question, pathways were categorized into groups. Since the aim was to find out which sort of pathways lead to benefits, the formation of groups began from this angle, but other distinctive attributes were also considered. To answer the second research question, we examined the patterns of pathways and found factors that influenced the benefits of PA. To verify our findings and explore more factors, we looked at students' pathways in detail.

Results

We found four distinctive groups of students' pathways (Fig. 2). The first group (1) represents students that improved their work after PA. We noticed some students (groups 2 and 3) not improving their work but still experiencing other benefits of PA. These students had different orientations to the original task regarding effort on original work. Since orientation to original work is a significant factor, two groups were formed: (2) students who did not improve their work, experienced other benefits of PA, and did not lack effort on original work; and (3) students who did not improve their work, experienced other benefits of PA, and lacked effort on original work. The last group (4) represents students that did not improve their work nor experience any other benefits of PA.

We will now introduce the pathways of four students, one from each group. Since all students' narratives, even within a group, have individual features, these must not be seen as an average case but as a distinct representative of the group.

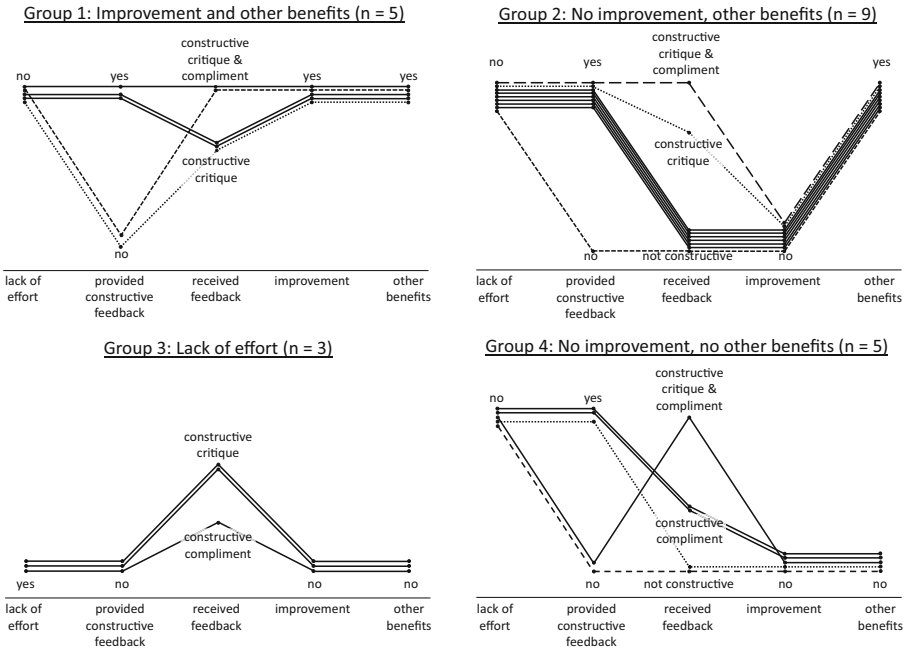

Figure 2 Students' PA pathways. Each line represents one student and his/her experience of this PA

Case Nea: Multiple Benefits

Nea is a student from group 1, which means that she experienced multiple benefits of PA (Fig. 3).

Reporting inquiry. Nea put effort into creating the lab report. In the interview afterwards, she stated that she "really enjoyed making the inquiry." The observer had written in the field notes that "Nea seems to do her best with the task and tries to follow the instructions. She got stuck with the formalities of reporting and does not quite understand the goal of the task. She seems to proceed, but slowly." Nea was one of many students who struggled with understanding the expectations of the lab report. Nevertheless, she enjoyed the task and put effort into it.

Providing feedback. When assessing her peer, Nea provided good-quality feedback, including both constructive critique and constructive compliment. The report that Nea assessed was rather good. Three out of four times, Nea marked correctly whether the assessee had achieved the requirements of the criteria, and in all four cases she provided constructive comments (Figs. 4 and 5).

Nea not only marked the smiley face but also specified her thoughts. Her comment communicates to the assessee that assessor had put thought into it, which increases the validity of the mark.

Here, Nea had noticed a shortage in assessee's work and specified which part of criteria needed completion. This gave assessee guidance how to complete his work.

Receiving feedback and making changes. Nea received good feedback with one constructive compliment and one constructive critique (Fig. 6).

The critique was valid; Nea had written unclearly about their measurements. The assessor had correctly marked the "serious face" and commented on the shortage of information.

After giving and receiving feedback, Nea reworked her report. She added a whole section to clarify how they had measured, which raised the quality of her work. In the interview, Nea explained her thoughts on receiving feedback:

> Interviewer: Ok, then. You had finished it [the report] and it was like this [gives a copy of the original report] and then you got your feedback. Here, you can look at it [gives a copy of the feedback]. So, do you remember the moment you got the feedback and read it, what did you think?
> Nea: Well, I thought that my text should have had that information and the feedback was good that I … like I was able to change my text with it.
> (a few turns later)

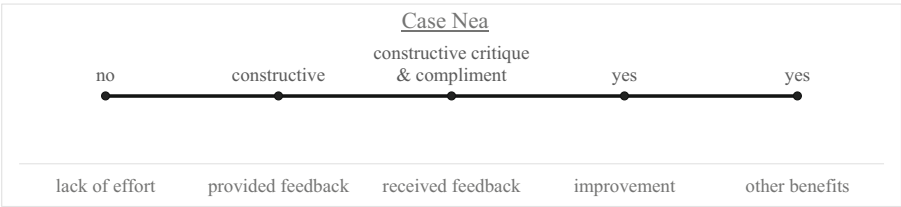

Figure 3 Nea's pathway through PA

Measurements are reasonable and clearly presented.				*Measurements were clearly in the text and seemed reasonable*

Figure 4 Example of a constructive compliment Nea provided

Interviewer: Yeah yeah, ok. So it was like … would you say that it was useful feedback?
Nea: Yes.
Interviewer: Ok. Was it easy for you to receive it or did it make you to feel …
Nea: It was easy!

Nea was receptive and grateful for the feedback, and she eagerly expressed that it was easy to receive. She wanted to use the feedback for her own benefit. This quality, being open to feedback and seeing it as a helping hand, is not self-evident, as we will see later in Milena's pathway. Nea had an open attitude toward feedback, which enabled her to improve her work.

Other benefits. In the interview, Nea stated that she considered all feedback useful. She mentioned that besides the critique, it was also helpful to know when something went well. Thus, she brought up another benefit besides improving her work, which is "confirmation," knowledge of something being properly done.

Nea's pathway is the one hoped for. She acted effectively both in the roles of assessor and assessee and benefited from PA. She had difficulties grasping the idea of conducting and reporting the inquiry, but it did not become a barrier to helping other students nor letting them help her.

Case Niko: Low-Quality Received Feedback

Many students made their inquiry carefully and provided good-quality feedback. Nevertheless, most of them did not improve their work. Niko was one of these group 2 students. He put effort into his work, but still ended up making no improvements after PA (Fig. 7).

Reporting inquiry and providing feedback. Niko worked hard to finish his lab report. The observer wrote in the field notes that Niko "wants to do his best and besides delivering a good report, he wants to learn and understand what he is doing." As assessor, Niko provided good-quality feedback. In three out of four criteria, he marked his peer's work correctly and in all those cases, he provided constructive written comments, of which two were constructive compliments and one constructive critique.

Niko put effort into the original work. He had one clear incomplete regarding error analysis, which could have been improved, but otherwise his work was of excellent quality.

Report shows, how speed is calculated from measurements. The calculation includes an equation, numbers and units.				*You could have added the equation to calculation*

Figure 5 Example of constructive critique Nea provided

Research plan and equipment have been clearly presented.				*You could have told more about materials that you used and how you made the measuring.*

Figure 6 Example of a constructive critique Nea received

Receiving feedback. The assessor of Niko's work focused on the wrong criteria and ended up providing poor feedback. She marked only one criterion correctly and did not include a written comment there. Three other criteria were marked wrong and had unconstructive comments since the focus was off (Fig. 8).

The criterion described the qualities of calculation. It was perfectly done in Niko's paper, but the assessor marked a serious face and commented only on the layout. There was a place for reinforcing feedback. Niko did not receive it and instead was led to focus on unimportant criteria.

Regarding the incompleteness of his error analysis, Niko did not get corrective feedback (Fig. 9).

The assessor had not noticed the incompleteness and instead marked a smiley face and complimented the work. In this case, Niko missed corrective feedback. He had another chance to notice the issue on the paper that he assessed since the error analysis was properly reflected there, but he missed that, too.

Changes. Niko made one small change to his lab report after the PA, but it was a superficial one and did not improve the level of his work. He changed two words in one sentence concerning materials. First, he had written, "Stopwatch and measuring tape *function* as materials," and he changed it to "Stopwatch and measuring tape *are needed* as materials." In the interview, Niko explained that this change might have made the sentence clearer, which shows his open attitude for making changes.

The feedback was not overly critical, but Niko took it seriously. Though two of four comments were positive, he gave more weight to the corrective feedback:

> Interviewer: Ok. Here's your work, this is the original one. Here's the feedback you got. So, when you got the feedback, what thoughts came up or how did you feel?
> Niko: Well, I agree that it was heavy to read because it was all written together, I could have written things separately to make it clearer.
> Interviewer: So there was critique and you agreed with it, so … there were quite many positive comments, too. Did you agree with them also?
> Niko: I guess so. (quietly)

Layout issues were the first thing Niko brought up. Though Niko's work was excellent, apart from the error analysis, the feedback was not reinforcing. Even though he was

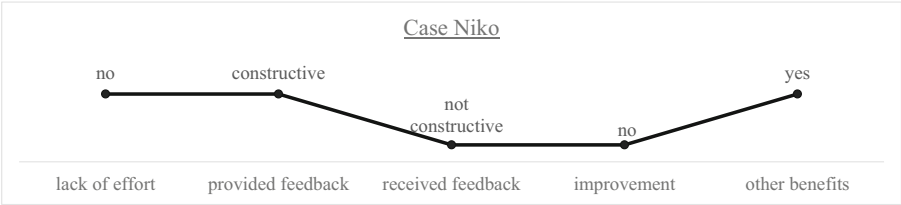

Figure 7 Niko's pathway through PA

Report shows, how speed is calculated from measurements. The calculation includes an equation, numbers and units.				*When all things were said in one paragraph, it became a little unclear to read but they are all there.*

Figure 8 Example of unconstructive critique Niko received

open to making changes to his work, he did not improve his work. The main factor was the quality of the feedback; it did not focus on the criteria.

Other benefits. In the interview, Niko stated that he considered PA useful and it gave him ideas for how to complete lab reports in the future. He said, "It [PA] was useful because when you assess others' work, they may have different aspects and you can notice it and fix yours." Here, he referred to layout issues, which were not central but not useless, either.

Niko's pathway shows the significance of received feedback. He put effort into his work and provided good feedback, but because of inadequate feedback received, he did not improve his work. Nevertheless, he determined the PA to be useful.

Case Ossi: Lack of Effort

Some students did not put effort into conducting the inquiry and creating lab report. In all these cases, the lack of engagement also resulted in providing low-quality feedback and making no improvements. Ossi's case is an example of the group 3 students (Fig. 10).

Reporting inquiry and providing feedback. Ossi did not put effort into conducting inquiry or making the lab report. The observer had written in the field notes, "Ossi was unmotivated and struggled with getting started with the task. After the first lesson, he had written just the headline and the purpose of the work. [During the lesson] he made several comments, saying that he does not care about physics nor physics grades."

Ossi provided low-quality feedback. On two criteria, he marked the serious face and commented "pretty ok," which is neither helpful nor encouraging. On one criteria he marked the sad face and commented "you could have told what you used," even though all the materials were listed in the report. It appears that Ossi did not do his best when assessing his peer's work. In addition, the observer noted that Ossi finished the task quickly, which supports the interpretation.

Receiving feedback and making changes. Ossi received constructive critique in all criteria, but did not use it to rework his report, which resulted in no changes. In the interview, he explained this as follows:

Interviewer: Was it useful, the feedback?
Ossi: No.
Interviewer: Why not?
Ossi: I would not have made changes anyway.
Interviewer: Was it because of the feedback that you did not make changes?
Ossi: No.

Error of measurements and results have been commented.				*These are quite clearly written!*

Figure 9 Example of unconstructive compliment Niko received

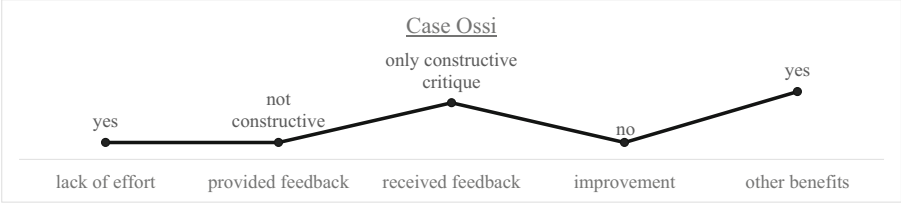

Figure 10 Ossi's pathway through PA

Interviewer: No. You had made your mind that it [report] was already good enough, had you?
Ossi: Yeah.
Interviewer: Could the feedback have been of the kind that you would have made some changes?
Ossi: No …
Interviewer: Do you think that teacher's feedback would have been more effective or just the same?
Ossi: Just the same.
(And a little later)
Interviewer: For me, this all seems useful feedback that you got, that you could have improved your work with this. Can you explain why you didn't want to make changes when you had the chance?
Ossi: I had just a little time left and I could not have finished it completely anyhow.

The main factor that kept Ossi from improving his work was not the quality of the peer feedback but his engagement with the work. First, he states that he had no intention of making changes and later explains that there was too much to do. This may be due to Ossi knowing that his report was far from the level he could reach. He may have improved his work without any feedback by only investing more time and effort, and peer feedback did not change it.

Other benefits. Despite the lack of effort, there was something positive. In the interview, Ossi expressed that he was reinforced by the feedback.

Interviewer: Here is the feedback you got. So, what did you think when you read it, or did you?
Ossi: Mmm. (Affirming)
Interviewer: How did you feel when you read it?
Ossi: That I have succeeded on some level at least.
Interviewer: Ok. Did it make you feel good?
Ossi: Yeah … possibly?

The words are modest, but in Ossi's case, they were significant. It was not often that he had implied there was something good in physics class. Though his report was not his best effort, he was enlightened that it was not all bad. The benefit of this experience did not result in an improved lab report, but it may have given him confidence on future projects.

Every student who had problems engaging with the original work had a similar pathway. The feedback they provided was not constructive and they did not improve their reports. Nevertheless, they experienced other benefits of PA. Based on this data, lack of engagement in the original work and in providing feedback is not a barrier to benefiting from PA.

Case Milena: No Use of Peer Assessment

The fourth group of students were students that did not experience any benefit from PA. They made no changes to their work nor brought up any other benefits. There are different reasons for these pathways. For example, the student who reciprocally assessed Ossi's poor work and received low-quality feedback from him could not think of any benefit of the PA. Logically, the combination of unconstructive feedback and less-than-superior work of the assessee led to experiencing no benefits from PA. There were also other explanations, which we will introduce in Milena's case (Fig. 11).

Reporting inquiry and providing feedback. Milena put effort into making the lab report. The observer wrote that she and her friends "concentrate on inquiry and try to accomplish it well. Occasionally they are more interested in how the rover moves than the inquiry, but still able stay concentrated." Milena did not provide constructive feedback. She gave only top marks, of which two were correctly marked but two were not. In the interview, she explained that her reason for giving positive feedback was not finding any big mistakes.

Receiving feedback, making changes, and other benefits. Milena received good feedback. She got two constructive compliments and one constructive critique. One criterion was wrongly marked as incomplete, but still had a good suggestion for how to improve her work. Despite receiving good feedback, Milena made no changes to her lab report nor brought up any benefits of PA in the interview. Based on her interview, one factor, defensiveness to peer feedback, became a barrier to benefiting from PA:

> Interviewer: You finished the work … and got feedback like this [hands out a copy], so how did you feel when you read it?
> Milena: I don't know, I guess I could have added the formula [talks about one constructive critique comment], but I had it there earlier in the text so she [assessor] could have paid more attention. And when here she wrote 'could there have been other errors' and 'was the measuring tape completely straight,' I remember having there that there could be some error, for example the timer started a little late and so on … Yes, there is the question, here. I thought it, but I

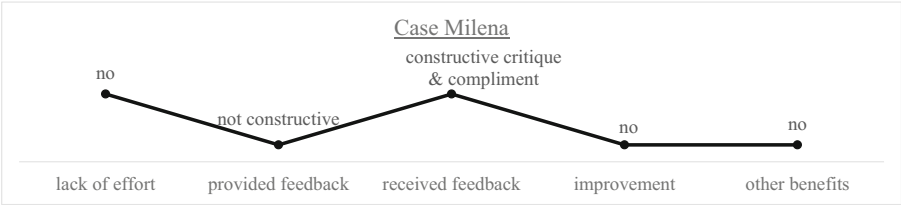

Figure 11 Milena's pathway through PA

didn't make the effort to write them all there. So, she [assessor] could have been more attentive there.

This is the first thought that Milena brings up about the feedback. She criticizes the feedback and explains what is wrong with it. She appears to see constructive criticism as offensive, not as a helping hand. In this case, the feedback was correct, but Milena did not use it for her benefit. Later, when asked, she did not mention any other benefits.

Interviewer: Would you say this feedback was useful?
Milena: Well I don't know … I don't know.

Above, Milena is reflecting on the matter, but does not come up with anything useful. This shows that even though a student originally puts effort into the work and receives good peer feedback, the benefits of PA are not guaranteed. Milena received useful feedback, but she found it offensive and ignored it.

Several pathways of this group indicated inadequate understanding of FA. This factor hindered their ability to benefit from PA. Defensiveness was one example, but also the mindset of "being done" led to similar results. These students felt that they had already worked enough or made a report of sufficient quality. They did not see assessment as something that is done for learning but as a procedure that is done in the end for others to judge their performance. These students did not improve their work, and, more interestingly, they also did not experience any other benefits.

Discussion

In this case study, we found four types of pathways through PA. They differed regarding the benefits of PA and effort on inquiry. Notably, the majority of students considered PA beneficial even if they suffered a lack of effort or received poor feedback (Fig. 2—all but one student from group 2 and all students from group 3). This demonstrates that benefits of PA are not solely due to received feedback and that the benefits cannot be reduced to improving one's work. This is in line with previous research, which has shown that PA may contribute to future success (Chetcuti & Cutajar, 2014; Tseng & Tsai, 2007), not just to work at hand.

We found three crucial factors for beneficial PA. These were students' own effort, receiving feedback that included constructive critique, and understanding the formative nature of assessment.

Putting effort into the original work was a necessary factor in order to improve lab reports after PA. While Nea put effort and time into her lab report and peer feedback, Ossi underachieved in both, using just a fraction of the time reserved for these activities. Both received feedback that included suggestions for how to improve their work. As a result, Nea made two improvements to her work, but Ossi continued with minimal effort and made no changes.

The lack of effort did not negate all benefits since everyone with low effort brought up at least one benefit of PA. Unlike Milena, who did not see any benefits of this particular PA, Ossi expressed that he was reinforced by the feedback. This does not necessarily mean that Milena has a less developed view of assessment; it is more

intimidating to receive critique of something that is one's best effort than something that could easily be improved. Hattie and Timperley (2007) state that in order to achieve his or her goal, students can sometimes lower that goal, which seems to be true in Ossi's case. He probably did not expect to receive very positive feedback since he had not pursued good-quality work. Still, he was enlightened to realize that he had achieved something. Since Milena had put effort into her work, it reflected her full capability, which made her more vulnerable to critique.

In this study, receiving feedback that included constructive critique was a crucial factor for improving one's own work. Figure 2 shows that all students who improved their lab reports had received constructive critical feedback (five students). Vice versa, critical feedback was relatively effective since most of the students that did not suffer from lack of effort improved their work if they received constructive critical feedback (five of eight students). This might be due to the novelty of conducting and reporting physics inquiry. Though the requirements of the work had been explained to students, the possibility of mistakes or misunderstandings was rather high. Uncertainty with the task may have kept the amount of critical comments low, but also lowered the threshold to react to critique.

The cases of Niko and Nea demonstrate the importance of constructive critique. Both put effort into their lab reports and were open to feedback. Nea received constructive feedback and improved her work, but Niko did not and, despite trying, failed to improve his lab report.

The results seem opposite to the findings of Tseng and Tsai (2007), who reported that corrective feedback is potentially harmful and therefore should be avoided. They do not report if 10th-grade students had any training or experience of PA. If not, that could partly explain the difference in results. Earlier in this study, when students were trained on PA, the researcher noticed students' resistance to critique. Afterwards, she planned a training session that specifically addressed the issue of receiving negative feedback and using it for one's own benefit. Another explanation for the opposite results obtained might be that this study observed the improvement in this particular task whereas Tseng and Tsai considered the improvement of subsequent work. This is in line with Hattie and Timperley (2007), who state that corrective feedback relating to some criteria is efficient in improving the work at hand, but it does not often generalize to other tasks. They also write that negative feedback can have a negative impact on subsequent performance and motivation.

In formative use of PA, a student's ability to deliver a proper grade is less important. In this study, receiving grades without justification did not seem to induce improvement since only the students who received written constructive critique improved their work. This is in line with Gielen et al. (2010), who found that justification was more important than the accuracy of the feedback. Though some students explained in their interview that they had noticed good qualities in their peer's work and learned something for the future, everyone stated that the improvements were attributable to received feedback. Though Anker-Hansen and Andrée (2019) found producing feedback being as helpful in promoting improvements, it does not seem to be the case here. In this study, receiving constructive critique appeared essential in order to improve a student's own work, which implies that learning to produce and receive feedback should be emphasized in PA training.

The third factor that influenced the benefits of PA was students' understanding of assessment. Seeing assessment as a helping hand and a learning tool instead of judgment facilitated receiving feedback and utilizing it. Milena and Nea both received similar feedback but responded to it differently. Nea was grateful for the critical feedback, which enabled her to improve her work, but Milena saw the same kind of feedback as criticism. This led her to reject it and not make improvements. One may speculate that Nea's ability to produce constructive critique was due to her relaxed attitude toward it. She pointed out the problems in her peer's work, while Milena did not. Milena gave full points "because nothing major was wrong." She may have supposed that critique would have offended her peer, like it offended her, and was not inclined to provide it. We suggest that internalizing FA should be a part of PA training. It is not enough to learn the cognitive skills needed in PA; students must also process the purpose of assessment in general.

The method "pathway analysis" appeared functional. We realize that the data of some stages—students' efforts, in particular—are less objective than others, but it seemed important to include this data in order to describe the whole pathway. The vast amount of data was almost obstructing at first, and to construe the cluster of students' pathways and find patterns, the information from each stage of the pathway was reduced to a minimum. The downside of gathering rich data was the moderate number of participants. In addition to the small quantity of participants, the specific context—one teacher, one school culture, and one culture—made the results ungeneralizable, which is a limitation of this study.

When student's motivation and understanding of assessment were adequate, PA provided multiple benefits for early-stage physics learners conducting physics inquiry. Inquiry is a core element of science and does not only advance understanding of scientific practices but also promotes growth in content knowledge (Marshall, Smart, & Alston, 2017). Using PA to let students help each other with inquiry has potential, but more research is needed on how the effect of PA develops in the long term when students continuously assess their peers during physics lessons. Do training and implementation of PA and gaining expertise in physics inquiry affect students' ability to benefit from PA?

Conclusions

We found that students who put effort into their own work, received constructive critique, and sufficiently understood the nature of FA were likely to improve their own work after reciprocal PA and also experience other benefits. PA is a complex intervention with many factors, and though its effects on students' learning have been studied, our analysis revealed new knowledge. In our study, students' age, the school subject, and the proportion of students who made changes to their work were similar to the study of Tsivitanidou et al. (2011). Our study contributed to this by describing individual students' pathways through PA and by finding factors that affected the benefits of PA. Of these, the quality of feedback and understanding of assessment need consideration when training and implementing PA. The significance of constructive critique differs from previous research (Tseng & Tsai, 2007) since in our study it appeared imperative in order to improve one's lab report. According to the findings,

peer-provided critique has potential to induce improvement in lab reports and should not be avoided but rather included in PA training. Utilizing critique is intertwined with understanding FA and should be discussed with students in order to increase the benefits of PA.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Anker-Hansen, J., & Andrée, M. (2019). Using and rejecting peer feedback in the science classroom: A study of students' negotiations on how to use peer feedback when designing experiments. *Research in Science & Technological Education, 37*(3), 346–365.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–144.

Chetcuti, D., & Cutajar, C. (2014). Implementing PA in a post-secondary (16–18) physics classroom. *International Journal of Science Education, 36*(18), 3101–3124.

Chong, I. (2017). How students' ability levels influence the relevance and accuracy of their feedback to peers: A case study. *Assessing Writing, 31*, 13–23.

Emden, M., & Sumfleth, E. (2016). Assessing students' experimentation processes in guided inquiry. *International Journal of Science and Mathematics Education, 14*(1), 29–54.

Finnish National Board of Education (FNBE). (1999). *Perusopetuksen oppilaan arvioinnin perusteet 1999* [The Fundamentals of Student Evaluation in Basic Education 1999]. Helsink, Finland: Opetushallitus.

Finnish National Board of Education (FNBE). (2014). *Perusopetuksen opetussuunnitelman perusteet 2014* [National Core Curriculum for Basic Education 2014]. Helsink, Finland: Opetushallitus.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304–315.

Gotwals, A. W., Philhower, J., Cisterna, D., & Bennett, S. (2015). Using video to examine formative assessment practices as measures of expertise for mathematics and science teachers. *International Journal of Science and Mathematics Education, 13*(2), 405–423.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hopkins, D. (2008). *A teacher's guide to classroom research* (4th ed.). Maidenhead: Open University Press.

Jindal-Snape, D., & Topping, K. (2010). Observational analysis within case study designs. In S. Rodrigues (Ed.), *Using analytical frameworks for classroom research: Collecting data and analysing narrative* (pp. 19–37). Abingdon: Routledge.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of PA. *Teaching in Higher Education, 11*(3), 279–290.

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science: An International Journal of the Learning Sciences, 40*(2), 257–275.

Marshall, J., Smart, J., & Alston, D. (2017). Inquiry-based instruction: A possible solution to improving student learning of both science concepts and scientific practices. *International Journal of Science and Mathematics Education, 15*(5), 777–796.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage Publications, Inc..

Ryan, J., & Gass, S. (2012). Stimulated recall. In R. Barnard & A. Burns (Eds.), *Researching language teacher cognition and practice: International case studies* (pp. 114–161). Bristol: Multilingual Matters. Retrieved from https://www.researchgate.net/publication/267573796_Ryan_J_Gass_S_2012_Stimulated_recall .

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks: Sage Publications, Inc..

Tasker, T., & Herrenkohl, L. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education, 27*(1), 35–59.

Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*(1), 20–27.

Topping, K. J. (2013). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 394–412). Thousand Oaks: Sage Publications, Inc..

Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49*, 1161–1174.

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

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*(4), 506–519.

Tsivitanidou, O. E., Zacharia, Z. C., Hovardas, T., & Nicolaou, A. (2012). Peer assessment among secondary school students: Introducing a peer feedback tool in the context of a computer supported inquiry learning environment in science. *The Journal of Computers in Mathematics and Science Teaching, 31*(4), 433–465.

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270–279.

Zhu, M., & Urhahne, D. (2014). Assessing teachers' judgements of students' academic motivation and emotions across two rating methods. *Educational Research and Evaluation, 20*(5), 411–427.