

JYU DISSERTATIONS 379

Laura Ketonen

Exploring Interconnections between Student Peer Assessment, Feedback Literacy and Agency



UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF EDUCATION AND
PSYCHOLOGY

JYU DISSERTATIONS 379

Laura Ketonen

Exploring Interconnections between Student Peer Assessment, Feedback Literacy and Agency

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi toukokuun 22. päivänä 2021 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education and Psychology of the University of Jyväskylä,
on May 22, 2021, at 12 o'clock noon.



JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2021

Editors

Pekka Mertala

Department of Teacher Education, University of Jyväskylä

Päivi Vuorio

Open Science Centre, University of Jyväskylä

Copyright © 2021, by University of Jyväskylä

Permanent link to this publication: <http://>

ISBN 978-951-39-8635-3 (PDF)

URN:ISBN:978-951-39-8635-3

ISSN 2489-9003

ABSTRACT

Ketonen, Laura

Exploring interconnections between student peer assessment, feedback literacy and agency

Jyväskylä: University of Jyväskylä, 2021, 92 p.

(JYU Dissertations

ISSN 2489-9003; 379)

ISBN 978-951-39-8635-3 (PDF)

This thesis aimed to advance the understanding of peer assessment, its dynamics, and its possibilities. The research involved multiple implementations of peer assessment in two lower secondary physics and chemistry classrooms in an urban school in Central Finland. The students ($n = 29$) were followed from the beginning of seventh grade to the middle of eighth grade. Data were collected using field notes, audio recordings, student interviews, students' written work, and written peer feedback. The qualitative data analyses were driven partly by data and partly by theory.

The first study examined the dynamics of a single peer assessment, asking who benefits from peer assessment and why. The analysis of individual students' pathways through peer assessment showed that receiving constructive critical feedback was beneficial for assessees. Otherwise, the students' role was significant because their engagement in the original task and their understanding of formative assessment influenced the benefits they experienced.

The second study explored students' feedback literacy (their understandings, capacities, and attitudes related to feedback) in the context of peer assessment, revealing a spectrum of skills that varied from neglecting feedback to actively seeking, processing, and using it. This variance must be accounted for when implementing peer assessment in the classroom. During the year of study, students developed their skills, implying that feedback literacy, so far a concern of higher education, can also be practiced at the secondary level.

The third study examined students' agency, specifically their capacity to act in the social context of the classroom during peer assessment. The analysis revealed several forms of agency and showed that students were unequally challenged by peer assessment. Certain forms of agency that were essential for productive peer assessment – such as judging others' work – were difficult for some students. The difficulties in exercising agency made students fall short in supporting one another's learning. Hence, besides a sense of responsibility, knowledge, and skills, students need agency to participate productively in peer assessment.

Keywords: peer assessment, formative assessment, feedback, feedback literacy, agency, secondary school, physics, chemistry, science

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Ketonen, Laura

Vertaisarviointi ja sen yhteys oppilaan palauteosaamiseen ja toimijuuteen

Jyväskylä: Jyväskylän yliopisto, 2021, 92 s.

(JYU Dissertations

ISSN 2489-9003; 379)

ISBN 978-951-39-8635-3 (PDF)

Tämän väitöstyön tarkoituksena oli ymmärtää vertaisarviointia ja sen mahdollisuuksia. Osallistujina oli kaksi perusopetuksen 7. luokan oppilasryhmää, jotka käyttivät formatiivista vertaisarviointia fysiikan ja kemian opinnoissaan. Oppilaita ($n = 29$) seurattiin seitsemännen luokan alusta kahdeksannen luokan puoliväliin. Tutkimusaineisto sisälsi oppituntien kenttämuistiinpanot ja ääninauhoitukset sekä oppilaiden yksilöhaastattelut, kirjalliset työt ja kirjalliset vertaispalautteet. Aineiston analyysi oli laadullinen ja pääsääntöisesti aineistolähtöinen.

Ensimmäisessä osatutkimuksessa syvennyttiin yhteen vertaisarviointiin ja tutkittiin, kuka hyötyi siitä ja miksi. Analyysi osoitti, että rakentava kriittinen palaute auttoi vastaanottajaa kehittämään työtään. Sen lisäksi vertaisarvioinnin hyödyllisyyteen vaikutti oppilaiden asennoituminen alkuperäiseen tehtävään sekä heidän ymmärryksensä formatiivisesta arvioinnista.

Toinen osatutkimus kartoitti oppilaiden palauteosaamista (palautteeseen liittyviä asenteita, taitoja ja ymmärrystä) vertaisarvioinnin kontekstissa. Oppilaiden palauteosaaminen oli hyvin eritasoista, ja se vaihteli palautteen vieroksumisesta sen aktiiviseen etsimiseen, pohtimiseen ja hyödyntämiseen. Vertaisarviointia käytettäessä on syytä huomioida palauteosaamisen erot. Oppilaiden palauteosaaminen kehittyi vuoden aikana, mikä osoittaa, että palauteosaamista kannattaa harjoitella jo peruskoulussa esimerkiksi juuri vertaisarviointia käyttämällä.

Kolmannessa osatutkimuksessa tutkittiin oppilaiden toimijuutta eli heidän kykyään toteuttaa vertaisarviointia luokan sosiaalisessa ympäristössä. Analyysi paljasti erilaisia toimijuuden muotoja ja osoitti, että näennäisesti yhdenmukainen vaatimus osallistua vertaisarviointiin ei ollutkaan käytännössä yhdenmukainen. Tietyt rakentavan toimijuuden muodot, kuten toisten auttaminen ja kritisointi, olivat osalle oppilaista tavanomaista toimintaa, kun taas toisilta ne vaativat oman totutun roolin ylittämistä. Jälkimmäisessä tapauksessa oppilaiden oli vaikeampi osallistua vertaisarviointiin rakentavasti. Tutkimuksen mukaan oppilaat tarvitsevat vertaisarvioinnin aikana tietojen, taitojen ja vastuuntunnon lisäksi toimijuutta.

Avainsanat: vertaisarviointi, formatiivinen arviointi, palaute, palauteosaaminen, toimijuus, peruskoulu, fysiikka, kemia

Author

Laura Ketonen
Department of Teacher Education
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
laura.k.ketonen@jyu.fi
<https://orcid.org/0000-0002-2821-0179>

Supervisors

Senior Lecturer, Docent Markus Hähkiöniemi
Department of Teacher Education
University of Jyväskylä, Finland

Professor Emeritus Jouni Viiri
Department of Teacher Education
University of Jyväskylä, Finland

Senior Lecturer Pasi Nieminen
Department of Teacher Education
University of Jyväskylä, Finland

Reviewers

Professor Päivi Atjonen
School of Educational Sciences and Psychology
University of Eastern Finland, Finland

Professor Anders Jönsson
Faculty of Education
Kristianstad University

Opponent

Professor Päivi Atjonen
School of Educational Sciences and Psychology
University of Eastern Finland, Finland

KIITOKSET/ACKNOWLEDGEMENTS

Olen siitä onnellisessa asemassa, että minua on töissä aina kannustettu kehittymään ja kokeilemaan. Olen erityisen kiitollinen koululleni ja sen matikistitiimille. Kiitos että valitsitte juuri minut töihin, otitte minut osaksi joukkoanne ja autoitte alkuun, kuuntelitte lennokkaita ideoitani ja vieläpä arvostitte niitä. Kiitos erityisesti Eeva, Leena, Nea, Tintti, Reijo, Rami, Markus ja Ilari. Meillä ei ole joutunut tekemään työtä yksin, vaan toisten tukeminen on ollut jokaisen kunnia-asia. Kiittäminen on myös koko koulun toimintakulttuuria, jossa kannustettiin ennakkoluulottomasti yrittämään ja erehtymään.

Tärkeimmät oppini arvioinnista olen saanut kaupungin arvioinnin kehittämistyöryhmässä. Kiitos erityisesti Pipsa, Jorma, Katri, Jarmo, Johnny, Esa, Tuija, Karri, Tiia, Risto, Pia, Tarja, Juhani ja vieraileva tähtemme Najat. Värikästä ja voimakastahtoista joukkoamme yhdisti intohimo arviointia kohtaan. Mieleeni ovat jääneet pitkien kokouspäivien tiukat väännöt ja opettajan näkökulmasta ylelliset lounaat, joihin kuului peräti kahvi ja jälkkäri ja joilla ei tarvinnut hätistellä lippiksiä naapuripöydän ruokailijoiden päästä.

Väitöskirjaprojektin realisoitumiseen tarvittiin edellä mainittujen altistavien tekijöiden lisäksi laukaiseva tekijä. Projekti ei olisi lähtenyt liikkeelle ilman ohjaajaani Jouni Viiriä, joka oli aina valmis keskustelemaan mahdollisista ja mahdottomista tutkimusideoista. Kiitos Jouni että mahdollistit tutkimukseni tekemisen ja houkuttelit minut töihin opettajankoulutuslaitokselle. Lämmin kiitos malu-ryhmälle, jonka osaksi heti tutkimusta aloittaessani pääsin. Kiitos Antti Lehtinen yliverstaistuesta, joka matkan edetessä on lähestynyt vertaistukea. Jos jäin jumiin, kävelin työpisteellesi ja neuvonpidon jälkeen tiesin lähes aina, miten edetä. Kiitos Anna-Leena Kähkönen neuvoista monenlaisten opetuksen ja tutkimuksen pulmien äärellä. Kiitos Joni Lämsä peesistä, jossa oli hyvä lähestyä väitöstutkimuksen maaliviivaa. Kiitos Sami Lehesvuori yhteistyöstä OPA -hankkeessa sekä ylipäätään siitä, että pääsin siihen töihin ja saatoin syventyä tutkimuksen maailmaan. Kiitos matkaseurasta Selma, Terhi, Kaisa, Sinikka, Anssi, Jarmo, Jenna, Jonathan ja Salla. Ja kiitos Josephine Moate avusta tiukoissa paikoissa otsikoiden ja sanavalintojen kanssa.

Haluan kiittää Jyväskylän opettajankoulutuslaitosta luottamuksesta ammattitaitooni ja hyvistä puitteista tehdä tutkimusta. Marja-Kristiina Lerkkanen, kiitos tuesta ja yhteistyöstä projektitutkimukseni tiimoilta. Sirpa Eskelä-Haapanen, en edes tiedä miltä kaikilta osin sinua on kiittäminen siitä, että olen löytänyt laitokselta oman paikan ja tutkimusaikaa, mutta ilman tukeasi tuskin olisin tässä tänään.

I want to thank my ESERA Summer School 2018 group – Justin Dillon, Berit Bungum and the other “Foxes.” Back then, I was a novice as a researcher, and I absorbed knowledge from you like a sponge. I owe my gratitude to Justin for your advice to stick with my original research idea and not “lose momentum.” That was probably the most important choice during this project.

Kiitos tietoturvasyistä anonymiksi jäävälle opettajalle siitä, että lähdit tekemään kanssani tutkimusta ja sen aikana huolehdit niistä langanpäistä, jotka

meinasivat minulta päästä purkautumaan. Kanssasi on ollut hyvä opettaa ja tutkia.

Erityinen kiitos kuuluu ohjaajilleni Markus Hähkiöniemelle ja Pasi Niemiselle. Kiitos panostuksesta, tuesta ja seurasta tämän projektin äärellä. Mieleeni nousee ensimmäisenä pitkät pohdinnat laitoksen ikkunattomissa luokkatiloissa, joissa paneuduimme milloin vertaiskoodaukseen, milloin alustavien analyysitulosten pähkäilyyn. Arvostan sitä, miten Markus pääohjaajanani jaksoit yhä uudelleen paneutua teksteihini, käydä niitä läpi tiheällä kammalla ja antaa palautetta monella eri tasolla. Näen, että tämä palaute ja sen sisäistäminen on merkittävin väitöstutkimuksen aikana tutkijan taitojani kehittänyt tekijä.

Lämmin kiitos työni esitarkastajille. Päivi Atjonen, en ihaile vain sitä tarkkaa ja asiantuntevaa palautetta, jonka tutkimuksestani annoit, vaan myös taidokkuutta, jolla olit sen muotoillut. Sen lisäksi että opin työni vahvuuksista, minun oli helppo motivoitua kehittämään työtäni kriittisten kommenttien perusteella. Pienillä vivahteilla loit kuvan siitä, että mahdollisimman korkealaatuinen työ oli meidän yhteinen intressimme. I would like to express my gratitude to another reviewer of this thesis, Anders Jönsson. Your feedback confirmed to me that my work was significant, not only in my own mind but also in the eyes of a feedback expert.

Ja viimeisempänä kiitos tärkeimmälle työyhteisölleni kuluneen vuoden aikana eli perheelleni. Vuosi sitten teimme kevään neljästään töitä kotioloissa, tämä lukuvuosi on edennyt hieman väljemmin. Kiitos että olette huomaavaisesti antaneet minulle tilaa ja rauhaa, vaikka olen vallannut omaksi työpisteekseni ruokapöydän. Kiitos seurasta työpäivän jälkeen kotisohvalla, laduilla ja poluilla. Joonas, kiitos että kannustat minua yrittämään ja iloitset saavutuksistani kuin omistasi. Yhdessä olemme olleet jakaneet kokemuksia ensin opiskelijoina, sitten opettajina ja nyt koulutusalan muissa tehtävissä. Yksi etappi on päättymässä, mutta seikkailu jatkuu.

Jyväskylä 12.4.2021

Laura Ketonen

LIST OF PUBLICATIONS

- Study 1 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, 1465–1484. <https://doi.org/10.1007/s10763-019-10030-3>
- Study 2 Ketonen, L., Nieminen, P, & Hähkiöniemi, M. (2020) The Development of secondary students' feedback literacy: Peer assessment as an intervention. *The Journal of Educational Research*, 113(6), 407–417. <https://doi.org/10.1080/00220671.2020.1835794>
- Study 3 Ketonen, L., Nieminen, P, & Hähkiöniemi, M. (under review). How do lower-secondary students exercise agency during formative peer assessment? *Educational Assessment*.

I was the first author of all three studies. I researched the literature, chose the methodology and theoretical framework, planned the interventions, gathered the data, and wrote the study. The coauthors contributed by providing feedback and perspectives on the method and my writing and by participating in peer coding and peer debriefings.

FIGURES

FIGURE 1:	The participants and functions of formative assessment (Black and Wiliam, 2009, p. 8).....	19
FIGURE 2:	Key strategies of formative peer assessment (adapted from Black and Wiliam, 2009)	20
FIGURE 3:	Intertwined features of feedback literacy (Carless & Boud, 2018, p. 1319)	28
FIGURE 4:	Relationship of peer assessment and feedback literacy	30
FIGURE 5:	Timeline of the training sessions, peer assessments (here abbreviated to "PA"), and interviews.....	40
FIGURE 6:	An example of students' pathway through peer assessment. This student lacked effort in the original work, he did not provide constructive feedback, he received only constructive critique about his work, he did not improve his work, and he experienced other benefits than improving work.....	46
FIGURE 7:	Students' pathways through peer assessment.....	49
FIGURE 8:	The forms of agency and the positions in which they were exercised	53

TABLES

TABLE 1:	Comparison of peer assessment studies in secondary school science classrooms	22
TABLE 2	The science and engineering practices of each inquiry task that were assessed by peers (abbreviated to "PA" in the table)	39
TABLE 3:	Descriptions of the tasks of and instructions for peer assessment (here abbreviated to "PA")	43
TABLE 4	A simplified criteria-based rubric for one category of feedback literacy. If a student, for example, did not make any changes to their work on seventh grade after peer assessment, but on eight grade made light changes, they moved from level 1 to level 2 in this category of feedback literacy.	47
TABLE 5	The criteria for the categories of feedback literacy	51

CONTENTS

ABSTRACT
TIIVISTELMÄ
ACKNOWLEDGEMENTS
FIGURES AND TABLES
CONTENTS

1	INTRODUCTION	13
2	THEORETICAL BACKGROUND	16
2.1	Interrelation of learning theories and assessment	16
2.2	Formative assessment is participative	18
2.3	Peer assessment.....	20
2.3.1	Defining peer assessment	20
2.3.2	Outcomes of peer assessment.....	21
2.3.3	Peer assessment requires training	24
2.4	Feedback as a shared responsibility	25
2.5	Feedback literacy enables productive feedback process.....	27
2.6	Feedback literacy and peer assessment support each other	30
2.7	Agency is inherent in formative assessment	32
3	STUDY AIM AND RESEARCH QUESTIONS.....	34
4	METHOD	35
4.1	Participants	35
4.2	Assessment in Finland	36
4.3	Researcher's role	37
4.4	Physics and chemistry studies	37
4.4.1	Peer assessment training.....	40
4.4.2	Peer assessments	42
4.5	Data and data collection	44
4.6	Analysis.....	45
5	OVERVIEW OF THE ORIGINAL STUDIES	48
5.1	Study 1: Pathways through Peer Assessment: Implementing Peer Assessment in a Lower Secondary Physics Classroom.....	48
5.2	Study 2: The Development of Secondary Students' Feedback Literacy: Formative Peer Assessment as an Intervention	50
5.3	Study 3: Exploring Students' Agency in Formative Peer Assessment	52
6	DISCUSSION	55
6.1	Owning the feedback process and practicing feedback skills.....	55

6.2	Taking on new constructive roles.....	58
6.3	Considerations of trustworthiness and ethical issues	59
6.4	Limitations and future directions.....	62
6.5	Final words	64
YHTEENVETO.....		66
REFERENCES.....		69
ORIGINAL STUDIES		

1 INTRODUCTION

The theory and practice of peer assessment have attracted growing national and international interest. In Finland, the National Core Curriculum for Basic Education (Finnish National Board of Education, 2014) obligated Finnish teachers to implement peer assessment in every subject for the first time. The change took place in 2016, when I was a subject teacher in a lower secondary school. As a proactive teacher, I was interested in implementing peer assessment in my classroom, but I was confused by two issues. First, I was not sure about the rationale behind peer assessment. Was the purpose to activate students, improve their learning results, or teach them life skills – or was it something else? Second, I could not easily find guidelines for how peer assessment should be implemented. I found it surprising that I, along with thousands of other teachers, had been instructed to adopt new teaching practices relying on only my professional intuition.

My bewilderment motivated me to study peer assessment. Researchers appeared to have a consensus on its usefulness grounded in the efficiency of formative assessment (Black & Wiliam, 1998a, 1998b). The first articles I read did not ease my confusion; the theoretical treatments (e.g., Topping, 1998; Topping, 2009; Topping, 2013) argued that peer assessment has multiple benefits, while the empirical ones investigated very specific aspects and often with modest or contradicting results (e.g., Anker-Hansen & Andrée, 2019; Chetcuti & Cutajar, 2014; Mok, 2011; Tsivitanidou et al., 2011; Tsivitanidou et al., 2012; Tsivitanidou, et al., 2018). In addition, my context was lower secondary education, but most of the research was on higher education, which is a very different environment, meaning the findings were not especially relevant to secondary education.

While undertaking this research, I began experimenting with peer assessment in my physics, chemistry, and mathematics classrooms. Students had questions regarding the practice, but we were able to discuss them, and the students accepted my instructions, which, in hindsight, were rather unpolished. However, the discussions were thought provoking for all of us, and without them, I would not have reached my current understanding of peer assessment. I found myself capable of implementing peer assessment, but because my students had

mixed feelings about the procedure, I was not entirely convinced of its practicality. Furthermore, apart from the constructive discussions with my students, the outcomes of peer assessment appeared modest, and the mystery of its utility remained unsolved.

This Ph.D. research arose from this confusion and my curiosity as my focus gradually shifted from the personal to the general. My principal aim was to explore peer assessment in depth and find out what happens in classrooms when peer assessment is implemented. In the classroom, the worlds of students and teachers are separate, and students learn to give the impression of involvement in classroom activities while hiding what they are really up to (Nuthall, 1995). As a teacher, I knew that I could only be partially aware of classroom occurrences. As a researcher, however, I could thoroughly explore all the discussions and products of peer assessment, interview students, put the pieces together, and discover the patterns, random occurrences, benefits, and constraints of peer assessment. I expected that an advanced understanding would lead me to develop much-needed recommendations for the practice of peer assessment. This was the first serious research project on peer assessment in Finnish basic education, and in addition to its contributions to international understanding of peer assessment, it would inform the national implementation.

Internationally, peer assessment has attracted a decent amount of research, but in this, certain perspectives have been emphasized. Psychometric qualities, such as validity and reliability of peer assessment, have received considerable attention (Panadero, 2016). Such qualities are crucial if peer assessment is used summatively, so that peers' judgements influence students' grades, but they are less significant if the aim is to advance students' learning or social and cultural features, such as collaboration. An inaccurate summative grade is inevitably a failure, but even the worst feedback comment can induce learning if it leads students to reflect the learning objectives, assessment criteria, and their work. Since the Finnish National Core Curricula for Basic Education (Finnish National Board of Education, 2020) instruct that peer assessment should be used only formatively, psychometric qualities are less essential. More vital is the knowledge of how peer assessment supports different students' learning and the classroom's learning culture. I agreed with the need to explore students' interactions during peer assessment and investigate not only the cognitive but also the sociocultural side of it (Panadero, 2016); thus, I wanted this Ph.D. research to fill this research gap. In addition, unlike most research on peer assessment, which has concentrated on higher education (Topping, 2017), I focused on lower secondary students. One more choice that I considered essential was investigating peer assessment, which was implemented according to the research with sufficient training to conduct peer assessment. However, I wanted to keep the intervention simple, so that it would be applicable to the teachers. From this basis, I planned and conducted my research.

In the following chapters, I share what I learned on my journey. I start by reviewing the evolution of learning theories, which partly explains the growing popularity of peer assessment. I continue by sharing my understanding of, and

defining the central concepts of the thesis. After presenting the aims of the study, I walk the reader through its implementation, present the main findings, and discuss their contributions and limitations. In the end, I return to the questions that provoked me to begin this study.

2 THEORETICAL BACKGROUND

2.1 Interrelation of learning theories and assessment

According to the online Cambridge Dictionary (n.d.), *assessment* means “the act of judging or deciding the amount, value, quality, or importance of something”. By nature, assessment is a value-laden activity (Boud & Falchikov, 2007). The choice of assessment objectives reflects what is considered central and valuable, and the choice of assessors reflects who is considered authorized and capable of assessing. The conceptions of knowledge and learning have influenced what is viewed as learning, what is considered worth learning, how assessment is conducted, and who is allowed to participate in assessment.

In traditional views of learning, the purpose of assessment is to confirm that transmitted knowledge has been received (Elwood, 2006; Gipps, 1999). Assessment appears straightforward because knowledge is considered objective, and assessment tasks are considered neutral and stable for all learners (Elwood & Murphy, 2015).

From the constructivist perspective, learning is a complex process, and assessment therefore aims to explore the quality and structure of students’ understanding (Gipps, 1999). For example, essays, projects, and concept maps can be used to encourage and evaluate deeper learning (Gipps, 1999; Won et al., 2017). According to constructivist views, teachers need to use assessment to elicit information on students’ understandings to lay the basis for formative assessment (Elwood, 2006; Elwood & Murphy, 2015). Constructivist approaches have been criticized for focusing on individuals and ignoring the problem of assessment being a value-laden social construct (Elwood & Murphy, 2015).

In socioculturalism, assessment encompasses both the process and the product, and it focuses on the social and cultural contexts of assessment and learning (Gipps, 1999). Assessment is not neutral (Elwood & Murphy, 2015), and therefore, students’ performances on assessment tasks can be understood by

looking “into [their] histories...not into their heads” (Elwood, 2006, p. 272). When participating in assessment, the assessee and assessors produce, reproduce, and transform society’s practices (Elwood & Murphy, 2015). From a sociocultural perspective, the focus of formative assessment is on the collective activity of understanding, which supports individuals’ learning and agency. Moreover, socioculturalism questions the assessment of individuals’ unsupported performance, as it considers the social environment and the tools it contains fundamental to learning. This idea is rooted in the work of Vygotsky (1978), who introduced the concept of students’ zone of proximal development, which is the level at which students can operate with the help of a more knowledgeable other but not yet on their own. From the Vygotskian perspective, assessing students’ best performance – that is, what they can do with help – is more valuable than assessing what they can do while unsupported (Gipps, 1999). Assessment methods that fit well with socioculturalism include portfolios, self-assessment, and peer assessment, but they should not be used rigidly, as Gipps (1999) argued: “Assessment within the framework of sociocultural theory is seen as interactive, dynamic, and collaborative. Rather than an external and formalized activity, assessment is integral to the teaching process and embedded in the social and cultural life of the classroom” (p. 378).

As explained above, the understanding of learning is transmitted to assessment practices. The impact is twofold, as assessment has a strong influence on learning (Biggs & Tang, 2011) because it speaks for what counts as learning and hence demands students’ attention (Bloxham & Boyd, 2007; Silseth & Gilde, 2019). Biggs and Tang (2011) called this the backwash effect of assessment. In an exam-dominated system, the effect can lead to the phenomenon of teaching to the test, but it can nevertheless have positive results if the assessment is built to support students’ learning. Hence, educators should carefully consider what they communicate through assessment and ensure that teaching and assessment are aligned (Biggs & Tang, 2011; Reeves, 2006).

Further, peer assessment aligns with the constructivist and sociocultural views of learning. Both value the active role that peer assessment gives to students. From the constructivist perspective, it promotes students’ own learning processes and constructions of knowledge, and can advance their understandings of the aims and processes of learning. From the sociocultural perspective, peer assessment is a part of social meaning making and negotiation. Social environment influences peer assessment, but correspondingly, peer assessment influences social environment.

In this thesis, I use a sociocultural framework to understand peer assessment, and I explore it by investigating a network of linked individuals and their interactions. The social plane of peer assessment is acknowledged throughout the thesis, but its emphasis is strongest in Study 3.

2.2 Formative assessment is participative

Assessment's main functions can be divided into the summative and the formative (Boud & Falchikov, 2006). Summative assessment judges learning results at the end of a learning unit, while formative assessment is used along the way to facilitate the learning process (Bennett, 2011; Boud & Falchikov, 2006). In the history of assessment, formative assessment is a relatively new practice. According to Gipps (1999), the first assessment tasks were summative, and they involved training the suitable and certifying the competent. The practice goes back 2,000 years to China, where it was used to select candidates for government service. Before summative assessment, family and patronage determined access to professions, but during the 19th century in Europe and America, examinations became a means of making education and career opportunities available to a wider group of people.

In 1967, Scriven (as cited in Bennett, 2011) made the distinction between assessment's summative and formative roles in the context of evaluating educational program, and in 1969, Bloom (as cited in Bennett, 2011) introduced the concept of students' formative evaluation. The most well-known definition of formative assessment may be Sadler's (1989), which was built on Ramaprasad's (1983) article on feedback from the perspective of management theory and outlined that the purpose of formative assessment is to close the gap between students' actual level and the target level. Sadler claimed that information becomes feedback only when it is used by the receiver.

The interest in formative assessment grew exponentially after Black and Wiliam (1998a, 1998b) emphasized the effect of formative assessment on students' achievement and presented impressive effect sizes for outcomes of formative interventions. Their results were later criticized (Bennett, 2011; Kingston & Nash, 2011), but the benefits of formative assessment have not been questioned.

The definitions of formative assessment have diverse emphases. Black and Wiliam (1998b) defined assessment as formative "when the information gathered in the assessment is actually used to adapt the teaching to meet student needs" (p. 140). By their definition, activities are not themselves formative but become so when the information they provide is actually used to advance learning. Cowie and Bell (1999) defined formative assessment as "the process used by teachers and students to recognize and respond to student learning and to enhance that learning, during the learning" (p. 101). Formative assessment is here defined as process, possibly to distinguish it from definitions that favored the taking of tests (Bennett, 2011). Equally, Black and Wiliam's definition went against this view by attributing formativity to its function rather than to an instrument. Clark (2012) emphasized the procedural nature of formative assessment and stated that its goal was for learners to self-regulate:

Formative assessment is not a measurement instrument; it is not designed to provide a summary of attainment at pre-determined intervals. Instead it is designed to continuously support teaching and learning by emphasizing the meta-cognitive skills and learning contexts required for [self-regulated learning]; planning, monitoring and a

critical yet non-judgmental reflection on learning, which both students and teachers use collaboratively to guide further learning and improve performance outcomes (p. 2017).

Clark's description of formative assessment paints a charming scene in which students and teachers work together for the sake of learning itself and not for extrinsic rewards, but how can such a situation be created? Black and Wiliam (2009) described the roles of formative assessment in processes of learning and teaching (see Figure 1) and argued that teachers are responsible for providing effective learning environments, while students are responsible for learning in those environments. Formative assessment occurs when students come to know the objectives of learning and criteria of success, learn what they already know, and figure out how they can advance their learning, all of which are accomplished through different activities, such as classroom discussions, teacher feedback, collaborative work, and peer and self-assessment.

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1 Clarifying learning intentions and criteria for success	2 Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3 Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4 Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	5 Activating students as the owners of their own learning	

FIGURE 1: The participants and functions of formative assessment (Black and Wiliam, 2009, p. 8)

Egelandstal and Riese (2020) criticized the conceptualization of formative assessment as a means of closing the gap. Using Gadamer and Dewey's concept of experience, they questioned the idea of learning as a linear process. They argued that because learners enter the classroom with their individual experiences and presuppositions, they interpret situations in unique ways, which results in unpredictable and unique outcomes. Hence, the thought of predestined learning objectives is unrealistic and, when used extensively in a controlling way, potentially indoctrinating. For assessments to be transparent, students need to know their learning objectives from the outset of a study unit (Boud, 2014). However, formative assessment should not be used to control learning or restrict its scope to predefined objectives but to support students' individual development and guide them in their learning trajectories (Silseth & Gilje, 2019). Clark's description of formative assessment also aligns with this view, as it emphasizes students' self-regulation and does not exclude any type of learning result.

2.3 Peer assessment

2.3.1 Defining peer assessment

Peer assessment can be used for summative and formative purposes (Topping, 2013). In this thesis, only the formative function is considered. Topping (1998) was the first researcher to develop and review a theory of peer assessment. He defined it as “an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of their peers of similar status” (Topping 1998, p. 250). Topping (2013) emphasized that both assessee and assessor are supposed to benefit from the process. Indeed, peer assessment is generally used reciprocally, meaning that students act as both the assessor and assessee and therefore experience the benefits of both roles.

Drawing on the key strategies for formative assessment (Black & Wiliam, 2009) presented in Figure 2, peer assessment is formative when its goal is to help students understand learning intentions (also called *learning objectives* or *goals*) and criteria for success and to activate them as instructional resources for one another. Teachers’ responsibility is to articulate that peer assessment aims to advance learning instead of measuring it and to build a learning environment and instructions that support that aim. Since students are supposed to learn from acting as both assessor and assessee, activating students as instructional resources for one another involves two distinct aspects – guiding students to be learning resources for their peers and guiding them to use their peers as learning resources (see Figure 2).

Where the learner is going	Where the learner is right now	How to get there
1 Student understands and shares learning intentions and criteria for success.	2 Student is an instructional resource for others.	
	3 Student uses others as an instructional resource.	

FIGURE 2: Key strategies of formative peer assessment (adapted from Black and Wiliam, 2009)

Topping’s (1998) definition only concerned the outcomes of subject learning, and a broader definition is needed to express that peer assessment affects not only the task at hand but also learning attitudes and learning strategies. Building on previous definitions (Carless & Boud, 2018; Cowie & Bell, 1999; Topping, 1998, 2013), formative peer assessment is defined in this thesis as a procedure in which students assess or are assessed by their peers with the intention that both assessees and assessors enhance their work or learning strategies in the process.

2.3.2 Outcomes of peer assessment

The majority of the research on peer assessment has concentrated on four main areas: validity and reliability, effects on learning, effects on self-regulated learning and metacognition, and the role of psychological and social factors (Panadero et al., 2018). However, the research is heavily focused on the first two areas, and the social factors are particularly under researched (Panadero, 2016; van Gennip et al., 2009). In addition, most peer assessment research focuses on higher education, leaving a gap at the secondary school level (Topping, 2013; van Zundert et al., 2010). As studies situated in secondary schools are fewer and tend to receive less attention, I will only discuss those when it comes to the outcomes of peer assessment.

The outcomes of peer assessment seem unpredictable, which is due to the various ways it can be implemented and the numerous uncontrollable factors involved in implementation. Topping (2013) listed 55 factors that should be reported when conducting a study on peer assessment. I demonstrate the inconsistency of outcomes of peer assessment by comparing four studies (see Table 1) that were more similar to this study than other offerings in the literature. All four focused on the secondary level and were conducted in science classrooms. In each, peer assessment was used formatively so that students were guided in assessing each other's work and were given opportunities to revise their own work afterward. Despite the studies' similarities, the percentage of students that revised their work varied from 0% to 91% among the studies. The findings underline the complexity of peer assessment, as even the easily measured variables are difficult to predict and explain. Therefore, in reviewing the studies' outcomes presented in Table 1, one must keep in mind that they speak more to the potential of peer assessment than what can be reliably expected from it.

TABLE 1: Comparison of peer assessment studies in secondary school science classrooms

Study	Student age/grade	N	Assessed task	Did the study observe changes or improvement in students work?	The percentage of students that changed or improved their work
Tsivitanidou et al., 2012	14/8	38	Designing a healthy pizza	Change	0%
Tsivitanidou et al., 2011	14/7	36	Designing a web portfolio for a CO ₂ -friendly house	Change	33%
Anker-Hansen & Andrée, 2019	14-16/8-9	98	Designing an experiment comparing the effect of two different breakfasts and exercises	Change	79%
Tsivitanidou, et al., 2018	unreported/11	22	Designing a model for color mixing light	Improvement	91 %

In terms of specific outcomes, peer assessment can promote subject skills. Lepak (2014) noticed that it advanced students' mathematical argumentation. Her study reported on a teacher's intervention in a class of low-achieving students. By using rubrics to evaluate their own and their peers' arguments, the students learned to write stronger and more coherent arguments, an outcome that continued after the intervention. Kim and Song (2006) reported that peer assessment improved eighth graders' scientific inquiry. In their study, students conducted open inquiries and wrote reports in small groups. In a peer review, each group presented their work and defended it while others acted as critics. The researchers noticed that both preparing to review and reviewing itself made students reflect on the inquiry and improve their interpretation and methods of experimentation.

Several studies focused on peer assessment's effects on specific subject skills, including writing. Gielen et al. (2010) investigated two classes of seventh graders who wrote essay drafts in their first language of Dutch, provided and received peer feedback, and then rewrote their drafts. The peer assessment improved students' writing and especially the justification of the feedback comments was found to have a positive effect on assessee's performance. Similarly, Kurihara (2017) investigated 35 17-18-year-old students studying English as second language, and Tsagari and Meletiadou (2015) investigated 60 13-14-year-old students. In both studies, the experimental groups that provided and received

peer feedback on essay drafts showed significant improvement compared to the control group that had only received teacher's feedback.

Peer assessment has been shown to be effective in improving programming skills as well. Wang et al. (2017) researched 166 ninth grade students in their programming classes. The students in experimental groups used online peer assessment as a part of their studies and outperformed the students in control groups that received only teacher's feedback. The experimental groups enjoyed better results in the programming project and on the final test.

As a final example, peer assessment can improve students' understanding of chemistry. Chang et al. (2009) focused on 271 seventh grade students learning about molecular models and chemical reactions with an animation tool. Students were assigned to three groups that received different kinds of treatment. The results showed that if students designed and interpreted their own animations and combined those efforts with peer assessment, the learning results were better than if they only designed and interpreted their own animations or viewed and interpreted their teachers'.

Apart from learning outcomes, peer assessment can promote positive attitudes. Johnson and Winterbottom (2011) reported that peer assessment among 28 15-16-year-old students in a girls-only class fueled their motivation to engage in science. Peer assessment has also been shown to heighten students' satisfaction with their studies. Hsia et al. (2016) compared 163 students (mean age of 14) in two settings. One group used a web-based and video-supported environment, and the other used web-based peer assessment, and the students that used peer assessment were more satisfied with their studies. The researchers also noticed that the students who used peer assessment had high levels of self-efficacy and motivation.

Additionally, peer assessment can develop students' metacognitive skills. Sadler (1989) proposed that it enables students to develop their self-monitoring skills by developing their self-assessment skills and strategies for closing the gap between their actual level and their goals. Several secondary education studies align with Sadler's claim. After using peer assessment, 13- to 14-year-old biology students' answers indicated that their understanding of the learning goals and their role in learning had advanced (Crane & Winterbottom, 2008). In Tasker and Herrenkohl's (2016) study, peer assessment combined with training and teacher's support advanced seventh grade science students' self-monitoring skills and made them more aware of the qualities of useful feedback. Students learned to provide more meaningful peer feedback and learned to evaluate the usefulness of the peer feedback they received.

Even though majority of the studies on peer assessment reported learning benefits, Le Hebel (2017) described a less successful trial. Her study examined 152 science students who assessed their peers and revised their work after assessment, and the author found that peer assessment was notably ineffective in identifying and correcting scientific misconceptions. Hence, peer assessment is not a solution to every situation.

Few secondary school studies highlighted complications with peer assessment, but some exceptions were found. Peterson and Irving (2008) pointed out that students disregard their peers' feedback and prefer their teacher's. Dolezal et al. (2018) noted that even though students were generally satisfied with peer assessment, some had experiences of being assessed unfairly. Mok (2011) described how some students questioned and were stressed about their ability to assess others. Tseng and Tsai (2007) found that critical and lecturing peer feedback negatively influences students' subsequent work.

The outcomes of peer assessment appear prevalently positive, but this must be taken with reservations, since the studies' interventions tended to be exceptionally well planned and conducted by well-informed people. More than proof of efficacy, the research on peer assessment requires a deep understanding of the phenomena that arise with it. Such an understanding would help explain the fluctuations in results and experiences, and assist in the development of policy and practice.

2.3.3 Peer assessment requires training

Researchers have a common understanding that peer assessment requires training for students (Gielen et al., 2010; Hovardas et al., 2014; Lu & Law, 2012; Panadero, 2016; Pandero et al. 2018; Topping, 2017; van Zundert et al., 2010). Training takes time and practice, as peer assessment involves multiple skills.

Sluijsmans (2002) identified three such skills: 1) defining the assessment criteria, 2) judging the performance of a peer, and 3) providing feedback for future learning. The students in Sluijsmans' study were in higher education, which explains the high requirements. Having students define assessment criteria is not mandatory for the implementation of peer assessment, but it does foster their ability to make judgements (Liu & Carless, 2006). The decision of whether to use student- or teacher-made criteria should be made with consideration of students' ages and subject skills. If students are just beginning to learn the concepts and character of the subject, they cannot be expected to develop solid criteria for success. In such cases, premade criteria can help them understand the expectations of a task (Panadero et al., 2013) and the requirements of a high-quality assessment (Gan & Hattie, 2014). Rubric use in peer assessment improves the accuracy and validity of feedback (Ashton & Davies, 2015; Panadero et al., 2013). One way or the other, the assessment criteria must be made familiar to students.

Judging the performance of a peer means comparing their work to the criteria and analyzing its strengths, weaknesses, and errors. Students' expertise improves the quality of their judgements. Falchikov and Goldfinch's (2001) review study found that the reliability of peer assessment in higher education is high overall but that its validity is higher in advanced courses than it is in introductory courses. Additionally, interpersonal issues can influence the reliability of peer assessment, such as worrying about hurting assessee's feelings (Cartney, 2010; Davis et al., 2007) or letting social relationships affect the feedback (Foley, 2013; Panadero et al., 2013). Training can improve students' psychological

safety during peer assessment (Panadero, 2016) and hence assist them in making solid judgements. Training is also relevant to teachers in this regard, as it allows them to establish an atmosphere in which challenging other people's ideas is appreciated (Tasker & Herrenkohl, 2016).

The third skill – providing feedback for future learning – relates to delivery and content of feedback. It involves evaluating which observations should be included in feedback, communicating the feedback clearly and in a proper tone, and providing guidance for improvement. Training can make students aware of the qualities of useful feedback (Tasker & Herrenkohl, 2016), thus promoting its provision. Additionally, guiding students' feedback provision with questions can improve the specificity of feedback (Gan & Hattie, 2014).

Sluismans' three skills pertain to assessors – that is, they are needed for the provision of good quality peer feedback. Hence, they only tell half the story. As peer assessment and feedback are two-way processes, the skills of assessees must also be considered. Carless and Boud's (2018) framework for feedback literacy describes assessee skills, although peer assessors need feedback literacy too (Han & Xu, 2019b). The features of feedback literacy—appreciating feedback, judging it, managing affect, and acting on it—are skills that enable the receiver to benefit from peer assessment, and these aspects should be included in peer assessment training. The features of feedback literacy and their relationship with peer assessment are discussed in more detail in Subsection 2.5 and Subsection 2.6.

Unfortunately, most of the studies on peer assessment do not describe the design of peer assessment training provided to participants. Such information could bring to light the varying outcomes of peer assessment and would assist in evaluating and developing training programs.

2.4 Feedback as a shared responsibility

Formative assessment has separate functions for teachers and students. For teachers, formative assessment provides information about students' understandings and skills and enables them to adapt the teaching according to students' needs (Black & William, 2018). For students, formative assessment provides information and support that promotes their learning. The information that supports students' learning is called *feedback* (Winstone & Carless, 2020).

Feedback has a strong influence on learning, but its impact can be either negative or positive (Hattie & Timperley, 2007). Hence, mastering this powerful but delicate tool is vital for teachers. Feedback can come in many forms, such as hints, prompts, and questions (Hafen et al., 2015). Multiple factors influence its effectiveness, including focus, form, timing, and context (Wisniewski et al., 2020). In their meta-analysis of feedback, Hattie and Timperley (2007) defined 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). They introduced two perspectives on efficient feedback. First, feedback should answer three questions: 1) What are my goals? 2) How am I doing in relation to

those goals? and 3) How can I proceed? Instances of feedback providing information about the first and third questions are sometimes called “feed up” and “feed forward,” respectively (Chong, 2020; Clark, 2012). However, in this study, these terms are not employed because the definition of *feedback* by Hattie and Timperley (2007) already entails both elements. Moreover, these terms can undervalue students’ agency and responsibility in the feedback process (Reimann et al., 2019). This is because they imply that the content of the feedback defines whether it functions as feed up, feedback, or feed forward, whereas the learner’s response also plays a role. For example, guidance can be interpreted as criticism, whereas knowledge about shortcomings can be viewed as guidance on how to proceed. In a similar vein, the terms “formative feedback” and “summative feedback” are used in some studies, but since sophisticated definitions of feedback, such as that of Hattie and Timperley (2007), entail both a formative function (how to proceed) and a summative function (how the learner is doing in relation to the goals), such terms blur as opposed to clarifying the conceptualization of feedback.

Second, according to Hattie and Timperley (2007), feedback can focus on four levels—the levels of task, process, self-regulation, and self. Task-level feedback is effective in assisting with specific work, but its restriction is that it does not generalize to different types of tasks. Feedback about process and self-regulation supports students’ learning in the longer term and enhances deeper learning. Feedback about self can even be harmful for learning because it passivates students by focusing on features that students cannot change and mediating an image of ability as a fixed property (Haimovitz & Dweck, 2017).

Although teachers are not the only source of feedback in Hattie and Timperley’s definition, the authors claim that teachers are responsible for ensuring its appropriate timing and focus. While the authors recognized students’ roles, they did so principally by framing them as objects of feedback. This is a traditional teacher-centered view of feedback as information; it is also referred to as the *old paradigm* of feedback (Chong, 2020; Nash & Winstone, 2017; Winstone et al., 2020).

As in contemporary learning theories, the conceptualizing of feedback has become more student centered. During the 2010s, interest began to focus on the receivers of feedback (e.g., Boud & Molloy, 2013; Carless & Boud, 2018; Dawson et al., 2019; Delva et al., 2013; Jonsson, 2013; Sutton, 2012; Wiliam, 2012; Winstone et al., 2017). It is now argued that teachers’ are not only responsible for providing feedback but also ensuring its reception and students’ utilization of it (Boud & Molloy, 2013). Students face barriers in that utilization, but having both teachers and students share the responsibility for feedback assists in overcoming those barriers (Winstone et al., 2017). To highlight students’ active role, feedback has been defined as “a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies” (Carless & Boud, 2018, p. 1315). According to that definition, feedback is not an episodic deliverance of piecemeal information but a process. It is led by assesses, and teachers facilitate the process and development of students’ skills. Such a

view is called the new paradigm of feedback (Chong, 2020; Nash & Winstone, 2017; Winstone et al., 2020). According to Sadler (1989), information becomes feedback only when it is used by the receiver. From this perspective, the quality of feedback is less important than the reactions it provokes in the receiver. Even poor feedback can initiate a process that benefits the learner, but feedback that is never read has no impact.

Naturally, both perspectives—the quality of feedback and its reception—are valid. Ideally, the teacher and student share responsibility for the feedback process, the feedback is constructive and timely, and the receiver is capable and willing to use it. Nevertheless, the focus of feedback research has shifted. As feedback was formerly understood as teachers' responsibility, it is now conceptualized as a process that primarily belongs to students (Dawson et al., 2019; Molloy et al., 2019; Winstone et al., 2020).

2.5 Feedback literacy enables productive feedback process

Students' feedback literacy refers to the skills they need in managing their feedback processes. Sutton (2012) conceptualized feedback literacy as the ability to read, interpret, and use written feedback. Carless and Boud (2018) built on his work and defined feedback literacy as “the understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies” (p. 1316). Feedback literacy emphasizes students' engagement with feedback and their active role in feedback processes.

Carless and Boud's framework (2018) presents feedback literacy as a composite of four features: appreciating feedback, making judgments, managing affect, and taking action. *Appreciating feedback* is about recognizing the significance of feedback for learning, as well as understanding that the feedback process requires students' active participation. It entails valuing feedback that comes not only from the teacher but also from other sources. Learning to appreciate feedback requires disrupting conceptions of teacher-centered summative feedback practices. *Making judgments* refers to students' ability to appraise the quality of own and others' work. These skills are needed and practiced in peer and self-assessment. The capability to make judgements involves learning the criteria and qualities of good work and understanding how to connect these qualities to specific aspects of a work. *Managing affect* refers to maintaining emotional balance when engaging with feedback. It is about dealing with personal emotions so that they do not disturb the feedback process. With a sufficient control of the emotions, students can strive for continuous improvement, have dialogues about feedback, and avoid defensiveness. These three features of feedback literacy (appreciating feedback, making judgements, and managing affect) are interrelated (Figure 3). For example, learning to understand that critical feedback aims to improve learners' performance (appreciating feedback) may assist in managing affect, and learning to manage affect may assist in participating in peer assessment (making judgements). When

successful, all three features give students more opportunities with the fourth feature, *taking action*. Taking action involves developing strategies for acting on feedback and understanding that using it requires recipients to act. Students should understand that taking action is the culminating point and the aim of the feedback process.

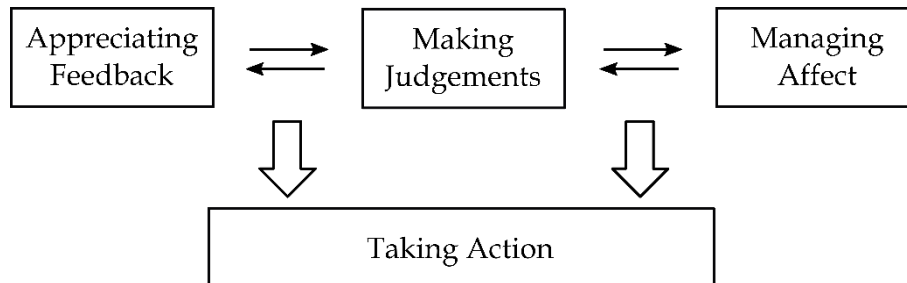


FIGURE 3: Intertwined features of feedback literacy (Carless & Boud, 2018, p. 1319)

These four features of feedback literacy describe students' cognitive, affective, and behavioral engagement with feedback. From a sociocultural perspective, feedback literacy does not only involve an engagement dimension but also a contextual and individual dimension (Chong, 2020). Contextual factors—such as the features of the feedback, materials, instructions, social relationships, and classroom roles—and individual factors—such as students' goals, attitudes, previous experiences, and academic abilities—influence the ways in which students engage feedback (Chong, 2020). Even students' goals, attitudes, and, to a certain extent, previous experiences are affected by their social environments. Thus, feedback literacy is not merely about individual attributes but is influenced by social context. Consequently, students' feedback literacy is rooted in individual and environmental development. Therefore, students' feedback literacy may vary from context to context.

Carless and Boud's (2018) work has inspired further research. Students' feedback literacy has been explored from a student perspective (Molloy et al., 2019). The authors identify seven groups of feedback literacy consisting of 31 categories of knowledge, capabilities, and skills, supplementing feedback literacy with new nuances. For example, they explicitly highlighted the ability to selectively accept and reject feedback, and they also underscored the understanding of expertise as a developing and unfixed feature. Feedback literacy has also been investigated in the context of an academic writing program (Hey-Cunningham et al., 2020). Here, the program for research students and supervisors entailed principles of feedback, literature exemplars, and peer and self-assessments, and it developed both students' and supervisors' feedback literacy. Especially, students learned what to do with feedback and how to use it efficiently. Han and Xu (2019b) explored how using teachers' feedback on feedback provided by peers influenced higher education students' feedback literacy. The students in the study developed their feedback literacy, but the intervention was more influential for the two motivated participants than it was

for the third participant with low motivation to engage in academic studies. Based on this, students' individual attributes play a role in the development of feedback literacy. Han and Xu (2019a) also investigated higher education students' profiles of feedback literacy and profiles' impact on students' engagement with feedback. The researchers found that the elements of students' feedback literacy were unbalanced, which reduced their capacity to engage with teachers' written corrective feedback. Students' feedback literacy was also situated so that their engagement with feedback depended on the current task and instructions and students' relating believes and motivation.

Giving students an active role in assessment and feedback processes supports the development of their feedback literacy (Carless & Winstone, 2020). This requires carefully designed feedback processes that include participative assessment practices, developing students' understanding of good quality, and planning the timing of feedback so that students can use it to advance their learning and work (Carless & Winstone, 2020). Practices that are considered to support students' feedback literacy are peer assessment (Carless & Boud, 2018; Chong, 2020), in which processes of providing and receiving feedback are both influential (Carless & Winstone, 2020), and the use of exemplars (Carless & Boud, 2018) – that is, samples of a work that represent dimensions of quality (Carless & Chan, 2017). The use of exemplars is especially efficient when accompanied with dialogues (Chong, 2019). Meta-dialogues about feedback are efficient for the development of students' feedback literacy (Carless & Boud, 2018), and they are an intrinsic part of both peer assessment and the use of exemplars.

Feedback literacy is reminiscent of another framework – that of assessment literacy. There are some conceptualizations of students' assessment literacy, meaning students' understanding of rules and standards of assessment in the educational context and their ability to use assessment tasks to monitor and advance their learning (Smith et al., 2013). However, most often, assessment literacy is used in relation to teachers – their knowledge about assessment, their conceptions and beliefs about assessment, their ability to make compromises in assessment between their beliefs and external factors, and awareness of their identity as assessors (Xu & Brown, 2016). Assessment literacy includes elements of feedback literacy, such as understandings of feedback and assessment purposes (Xu & Brown, 2016). The main difference between the concepts is that assessment literacy focuses on assessment in the educational context, whereas feedback literacy is context free. The ability to seek, process, and use feedback is needed not only at school but also at work and in private life. In the educational context, feedback literacy is a sub-feature of assessment literacy, but while assessment literacy is rather irrelevant outside the educational context, feedback literacy is an important life skill.

Judging by researchers' eagerness to build on Carless and Boud's framework, feedback literacy is a valuable concept in contemporary discussions about the transformation of feedback practices. However, empirical research on feedback literacy and its development is still scant, even more so outside the context of higher education. Therefore, new approaches to the topic are needed.

2.6 Feedback literacy and peer assessment support each other

Peer assessment and feedback literacy are interrelated (see Figure 4). Productive participation in peer assessment requires feedback literacy (Han & Xu, 2019b), and peer assessment is a platform to practice and advance it. If students have sufficient feedback literacy skills, peer assessment is more likely to be useful and comfortable to them. Since teachers are more likely to use peer assessment after positive experiences with it (Pandero & Brown, 2017), the student groups with higher feedback literacy skills are more likely to have the opportunity to use it more frequently, and thus, they will further develop their feedback literacy. Before implementation of peer assessment, students should be trained to use it (Gielen et al., 2010; Hovardas et al., 2014; Lu & Law, 2012; Topping, 2009; van Zundert et al., 2010); many aspects of the training relate to feedback literacy skills. Without a sufficient level of feedback literacy, students cannot provide and use feedback, and the peer assessment is more likely to malfunction and be rejected.

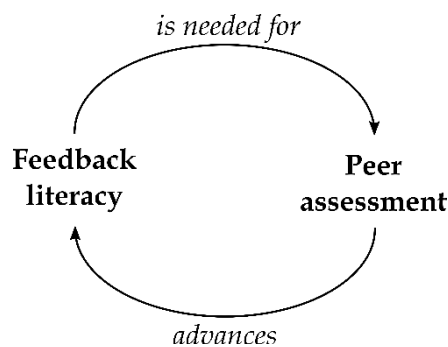


FIGURE 4: Relationship of peer assessment and feedback literacy

Prior research on peer assessment recognizes features of feedback literacy, but it does not generally mention it. Next, I introduce research on how each of the four features of feedback literacy (appreciating feedback, making judgements, managing affect, and taking action) (Carless & Boud, 2018) is necessary in peer assessment, as well as research that shows how peer assessment can help to develop these features.

Appreciating feedback. If students do not understand that formative peer assessment is supposed to advance learning, they may perform it unproductively; for example, they might provide only positive, superficial feedback (Tasker & Herrenkohl, 2016) or let their relationships with assessees influence the feedback they provide (Foley, 2013; Panadero et al., 2013) – often referred to as *friendship marking*. Peer assessment provides opportunities to discuss the differences between formative and summative assessment (Davis et al., 2007). Moreover, peer assessment training provides a context for discussions and reflection that can change students’ peer feedback to make it more substantial and develop their appreciation of critical and guiding feedback (Tasker & Herrenkohl, 2016). It is essential for productive peer assessment that students appreciate feedback not

only from their teacher but also from their peers. Although students often disregard and undervalue peers' feedback (Foley, 2013; Panadero, 2016), the use of peer assessment can increase their appreciation of their peers as a source of feedback (Crane & Winterbottom, 2008).

Making judgements. A requirement of productive peer assessment is that students make judgements about the quality of own and other students' work (Carless & Boud, 2018), as well as about received feedback (Molloy et al., 2019). Peer feedback is more beneficial if students interpret it critically, but they do not necessarily have skills to do this (To & Panadero, 2019). Students' engagement in peer assessment can be strengthened by instructing them to be active and critical as assesses and evaluate the received peer feedback (Minjeong, 2009). In addition, students need to understand the assessment criteria to be able to judge their peers' work (Cartney, 2010; Foley, 2013; Panadero et al., 2013). Students are more comfortable with peer assessment if they share an understanding of the criteria (Panadero et al., 2013), but also, engaging in peer assessment advances students' understanding of the criteria (Anker-Hansen & Andrée, 2019; Black & Wiliam, 2018) and develops their capability to judge their peers' work (Han & Xu, 2019b).

Managing affect. Peer assessment is emotionally challenging for students, and it can raise negative feelings (Cartney, 2010; Panadero, 2016). The feedback can make assesses defensive (Anker-Hansen & Andrée, 2019; Tasker & Herrenkohl, 2016), and assessors can worry that their feedback will raise negative feelings in assesses (Cartney, 2010; Davis et al., 2007; To & Panadero, 2019). However, since peer assessment is an interactive process, affective issues are inseparable from it. Therefore, for productive peer assessment, it is vital that students manage their emotions, and to do that, they need support on emotional aspects relating to both the assesses' and assessors' roles (Cartney, 2010). Peer assessment assists with affective issues because it supports the psychological safety in the classroom (van Gennip et al., 2010); as a result, students feel more comfortable asking for help and sharing their thoughts. Advancement of a culture in which sharing ideas and using others' help is common again supports the practice of peer assessment.

Taking action. Students can be reluctant to accept critical peer feedback and often do not use it to revise their work (Anker-Hansen & Andrée, 2019; Tsivitanidou et al., 2011; Tsivitanidou et al., 2012). Reluctance can derive from insufficient feedback literacy (Carless & Boud 2018), for example, from confusion about assessment criteria (Tsivitanidou et al., 2011). Peer assessment is useful in addressing and discussing these issues, and it can encourage students to act on feedback (Jonsson, 2012), particularly when it is combined with profound conversations about assessment (Cartney, 2010).

The research on feedback literacy has not focused on secondary education. The growing amount of research on students' feedback literacy and its facilitation has so far focused on higher education. As students' feedback literacy also appears to promote productive practices at secondary school, it should be intentionally nurtured and researched in that context as well.

2.7 Agency is necessary for formative assessment

As conceptions of learning have shifted from teacher centered to more student centered, students' agency has become essential. Students are encouraged to be agents of their own learning. Since teaching and assessment should be aligned, agency is also a necessary ingredient in assessment (Charteris & Thomas, 2016; Boud & Falchikov, 2006; Harris et al. 2018).

Agency is a challenging concept. It has various definitions and emphases, and the term is often used without definition (Eteläpelto et al., 2011; Rajala, 2016; Matusov et al., 2016). In psychology, agency is considered an individual's feature (Eteläpelto et al., 2013; Rajala, 2016) related to self-efficacy (Bandura, 2011). In socioculturalism, the sociocultural context is viewed as shaping individual agency. Earlier sociocultural approaches emphasized the influence of social context over the individual (Eteläpelto, 2013; Rajala, 2016), but more recent approaches regard the relationship as dialectic, meaning that the social context supports and constrains an individual's agency, but on the other hand, individuals reproduce and transform the environment (Rajala, 2016).

In this study, I take the latter stand and consider agency as an interplay of individuals and their environment that can be defined as "socio-culturally mediated capacity to act" (Ahearn, 2001, p. 118). According to this definition, agency is not merely the attribute of individuals but is rather based in context and achieved through engagement (Biesta & Tedder, 2007). Individuals develop patterns of agency that create expectations for their participation in the group (Gresalfi et al., 2009), and they vary depending on the context. Therefore, a person can exercise a certain kind of agency in one situation but not in another. When I analyze students' agency in this study, I use the following definition, which entails both the individual and the social context:

Individual's agency refers to the way in which he or she acts, or refrains from acting, and to the way in which her or his action contributes to the joint action of the group in which he or she is participating.¹

Three temporal elements—the past, present, and future—influence how an individual's agency is formed in a particular situation (Emirbayer & Mische, 1998). First, agency is built on experience. This means that a person's prior experiences influence the options they have and choices they make. Second, the outcomes a person imagines influences the action of their agency. Third, agency can only be enacted in the present. This means that a person must evaluate the demands and dilemmas of a present situation and act on them. Even though these three elements may restrict individuals' agency, one never entirely lacks agency but can, at a minimum, exercise it through compliance or resistance (Gresalfi et al., 2009).

In educational research, agency is mainly referred to as a positive thing, but it cannot be specified as good or bad (Rajala, 2016). It is exercised in many ways,

¹ Gresalfi et al. 2009, p. 53

including individuals' self-regulating, seeking help, setting goals (Harris et al., 2018), and transcending established patterns (Matusov, 2011). It does not necessarily require the bringing about of change (Biesta & Teddler, 2007) but can be exercised through resistance. Assessment resistance is an example of such agency (Harris et al., 2018). An even more contradictory form is *teacher-pleasing authorship* (Matusov, 2011), which signifies students' capability to anticipate and follow their teachers' desires and suppress their intentions. Such agency is encouraged in many school systems, but it is problematic. Similarly, the agency of resisting authorship cannot be categorized as simply right or wrong; instead, resistance reflects students' interpretations of the current situation and their position in it (Matusov et al., 2016; Rainio, 2008).

Nurturing students' agency by sharing with them the responsibility of assessment is one rationale for using peer assessment (Liu & Carless, 2006; Panadero, 2016; Topping, 2009). However, this agentic position can conflict with schools' other assessment practices. For example, technology-enhanced feedback that appears significant to students can be one-directional information provided by teachers that does not give students an active role (Oinas, 2020). As another example, in student-teacher-parent discussions that are intended as participative, teachers tend to share information about students' learning and behavior, and students struggle to participate actively (Luukkonen, 2020). Since schools' assessment and feedback practices comprise an incoherent mixture of teacher- and student-centered approaches, participating actively in peer assessment may challenge the students. There is no guarantee that they will embrace their active role. They may question their ability as assessors (Mok, 2011), resist their peers' feedback (Foley, 2013; Panadero, 2016), and worry about the effects of peer assessment on their social relationships (Harris & Brown, 2013). A productive and constructive social interaction, such as peer assessment, cannot be expected to emerge by itself, as it requires systematic work (Rajala, 2016).

The reciprocal seeking and providing of support is called "relational agency" (Edwards, 2009, p. 203), and it equates to the aspect of formative assessment (and peer assessment) that Black and Wiliam (2009) called "activating students as learning resources for each other" (p.8). The productivity of peer assessment is, to a large extent, dependent on students' relational agency. From this perspective, research on students' agency during peer assessment is pivotal. However, studies on assessment and agency are scarce (Nieminen & Tuohilampi, 2020), and even though students' considerations and experiences of peer assessment have been investigated (Panadero, 2016), little is known about the forms of agency that students actually exercise during formative peer assessment.

3 STUDY AIM AND RESEARCH QUESTIONS

This study's aim is to contribute to an understanding of how students act and develop as assessees and assessors when they train for and repeatedly practice peer assessment. Three individual studies were conducted, each examining the data of 1.5 years long peer assessment intervention through a different lens.

Study 1 investigated in detail a single peer assessment intervention and the pathways each student through it. The specific research questions were: 1) what kind of pathways do students take when peer assessment is implemented in a lower secondary school physics classroom and 2) which factors advance or reduce the benefits of peer assessment?

Study 2 investigated the skills and development of students' feedback literacy during a year of repeated implementations of peer assessment. The research questions were: 1) what kind of feedback literacy skills do students have in the context of reciprocal formative peer assessment and 2) how does students' feedback literacy develop throughout a year of using peer assessment?

Study 3 observed peer assessment through a sociocultural lens and examined students' agency during six separate peer assessments. The research questions were: 1) what forms of agency do students exercise during formative peer assessment and 2) How do students exercise agency in the different positions, with respect to other students, that peer assessment offers them?

4 METHOD

Since this thesis aimed to explore the implementation of peer assessment and provide new insights into it, I chose a case study design. Case studies are used to understand phenomena, clarifying how and why things happen (Thomas, 2021). As case studies typically use various data sources, such as observations, interviews, archives, questionnaires, and documents (Rowley, 2002; Thomas, 2021), I collected a rich data set that formed a comprehensive picture of peer assessment. The strength of case studies is that they allow a phenomenon to be researched in its real context (Rowley, 2002; Yin, 2009; Thomas, 2021), which I considered essential to this research. Without context—school, students, curricula, tasks—peer assessment loses its essence. The specifics of the case study must be reported carefully (Rowley, 2002), and the results must be interpreted in their context. Therefore, in the following subsections, I describe the study's context and course in detail, and in the appendices, I supplement this description with the materials used in the interventions.

4.1 Participants

This study was carried out in an urban school in Central Finland in 2018–2019. I conducted the interventions in cooperation with an experienced subject teacher and her two science classes of 15 and 16 students, which had roughly equal shares of boys and girls. The study started at the beginning of the students' seventh grade (when students' mean age was 13) and lasted until the middle of eighth grade. One student from each class declined to participate, bringing the total number of participants to 29. I asked the teacher to join the study because of her well-organized and thoughtful working style, which enabled the interventions' careful co-planning. In addition, we had experience of co-teaching together, and based on that background, I knew that we could naturally share responsibility in the classroom.

In Finland, students study general science in grades 1–6 with a class teacher, and this education includes biology, geography, physics, chemistry, and health. In seventh grade, they begin physics and chemistry studies with a subject teacher. Nevertheless, few participants recalled doing science inquiry activities, and the activities they could recall were cookbook experiments.

In grades 7–9, the school's students were appointed to classes of approximately 24 students. However, for physics and chemistry, two or three classes were mixed and shared with smaller classes to ensure students' safety during scientific experiments. Hence, at the beginning of the intervention, the students did not know each other well, though most had at least one friend in the class.

4.2 Assessment in Finland

The National Core Curriculum (Finnish National Board of Education, 2014, 2020) guides the assessment of basic education in Finland. There are no national exams or external control systems. The National Core Curriculum permits teachers to plan, implement, and assess its objectives as they see fit. For example, summative assessment data may consist of exams, tests, projects, investigations, and classroom interactions. No element is mandatory, but teachers must give students several ways of demonstrating their skills and knowledge. Finland has a strong tradition of summative assessment, and a summative view of assessment is still dominant (Atjonen et al., 2019). The concept of formative assessment was first introduced as an addition to National Core Curriculum in 1999, and National Core Curriculum of 2014 was the first to emphasize formative assessment.

According to the Curriculum (Finnish National Board of Education, 2020), students are supposed to receive feedback throughout their learning process. The National Core Curriculum of 2014 is based on the conception of students being active in their learning. With assessment, the curriculum pursues students' activity by guiding the teachers to ensure that students know the learning objectives and assessment criteria, and guiding teachers to implement self- and peer assessment with students. The use of formative peer and self-assessment is mandatory in all school subjects. Though the curriculum was released in 2014, change was gradual, and this study's participants only moved to it in fifth grade, two years before this study was conducted. However, as teaching practices change slowly, this study's participants had little experience with peer assessment. About half recalled a teacher implementing peer assessment in elementary school, but none reported regular experience with it.

4.3 The researcher's role

Since the aim of this study is to examine what can be realistically achieved by using peer assessment, it was naturalistically organized. My years of experience teaching in several lower secondary schools constituted a prolonged engagement (Lincoln & Cuba, 1985), and I had insider status at the school where the research was conducted. I had previous experience of co-teaching with the teacher that joined the study, and that made our cooperation natural. This opened the unique possibility of arranging an intervention and collecting the data with minimal disruption to normal classroom practices.

I planned the training sessions and peer assessments with the teacher. I made preliminary plans and adjusted them according to the teacher's preferences and the groups' weekly schedules. After lessons, the teacher and I shared our observations and considered whether any adjustments were needed.

I participated in most of the physics and chemistry lessons as an observer and assistant teacher, and I helped the teacher organize the training sessions and implement the peer assessments. I was presented to the students as a former science teacher at the school and as someone who was now working at the university of Jyväskylä as a researcher, and I was also described as an assistant teacher that could be asked for help. My role in the class soon became natural, and as of the first lesson I participated in, the students did not hesitate to ask for my help or share their thoughts. I avoided valuing or guiding their behavior because I did not want them to show me only their good sides. I emphasized that in my role as a researcher, I wanted to know what they thought about anything, not just positive things.

Conducting the study would not have been possible without the knowledge of the school. Unobtrusively embedding peer assessment into the curriculum, acting as an assistant teacher, organizing interviews, and reacting to unpredictable changes, such as the appearance of substitute teachers, was only possible because of my insider status. For example, arranging interviews required contacting multiple teachers, navigating the school, and finding students from different classrooms, and I also had to find free spaces in a packed school to conduct the interviews, including a bomb shelter, a locked staircase, a lunchroom, storerooms, and empty classrooms.

4.4 Physics and chemistry studies

In Finland, the National Core Curriculum (2014) prescribes the content and objectives of learning, but individual school districts decide how those are distributed during the grades 1-2, 3-6, and 7-9. In the district where this study was conducted, the seventh grade physics curriculum included the basics of inquiry, mechanics, and structures of the universe. Seventh grade chemistry was comprised of the basics of inquiry, including lab skills, the basics of properties,

and the structure of substances. The eighth grade physics curriculum included the basics of electricity and thermodynamics. The physics and chemistry courses each included approximately 18 1.5 hour classes for seventh grade and 27 1.5 hour classes for eighth grade.

The lessons took place in a science class that was adequately equipped for conducting simple scientific experiments. While completing the learning module, students received physics and chemistry textbooks that they returned afterward. The science teacher printed a homemade notebook for each student that contained written tasks, inquiry instructions, and theory.

Almost every lesson entailed student-centered inquiry-based activities that were conducted in groups of 2–4 students. Inquiry was used to support the learning of science concepts and laws, but learning to conduct and understand scientific inquiry was itself an important aim. Each peer assessment followed one of the varying inquiry tasks. These tasks are described immediately below, and the science and engineering practices they entailed are presented in Table 2. The peer assessment criteria for each task is presented in this study's appendices.

Task of Peer Assessment 1: Planning a rover. This was an engineering project with the goal of building a rover that could move on its own. The students could use any available material, such as straws, paper, cardboard boxes, plastic bottles, balloons, wooden sticks, rubber bands, and tape. The teacher provided an example picture of a rover to demonstrate what a drawn model could look like. The students could apply that model if they wished, but most groups decided to develop their own. The students were instructed to plan the rover in small groups, draw a model of it, and later build it. Planning and building the rover took approximately three 45 minutes lessons.

Task of Peer Assessment 2: Speed of the rover. After building the rover, students continued in the same groups and determined its speed. They were expected to measure an optional distance that their rover moved and the amount of time the movement took, calculate the speed, and write a lab report about their investigation. The instructions did not provide guidance for accomplishing the task, but the teacher supported the groups that needed help. While the inquiry was conducted in groups, everyone wrote their own lab report.

Tasks of Peer Assessment 3: Simple chemistry experiments. These tasks were chemistry experiments with straightforward instructions and questions. Each group was provided with the equipment and materials they needed for the tasks. The tasks were 1) measuring how much time it takes for water to boil when heated with a Bunsen burner, 2) observing the chemical reaction of burning magnesium, 3) testing how different substances dissolve in water, oil, and alcohol, and 4) investigating how a dissolution process influences the joined masses of the substances.

TABLE 2 The science and engineering practices of each inquiry task that were assessed by peers (abbreviated to “PA” in the table)

	Inquiry task of PA1	Inquiry task of PA2	Inquiry tasks of PA3	Inquiry task of PA4	Inquiry task of PA5
Asking questions and defining problems	-	-	-	-	-
Developing and using models	yes	-	-	-	-
Planning and carrying out investigations	-	yes	yes	yes	yes
Analyzing and interpreting data	-	yes	yes	-	yes
Using mathematics and computational thinking	-	yes	-	-	yes
Constructing explanations and designing solutions	yes	yes	-	-	yes
Engaging in arguments from evidence	during PA	during PA	-	during PA	during PA
Obtaining, evaluating, and communicating information	-	-	-	-	-

Task of Peer Assessment 4: Separation puzzle. The task was to separate a mixture of iron powder, salt, sand, and sawdust with the equipment available. Individually, the students made plans about the separation process. Later, they discussed their plans in groups, elaborated on them, and put them into action. The teacher gave each group a mixture that contained one spoonful of each substance. In the end, the outcomes of the groups’ separation attempts were examined with the whole class, and variances in results and approaches were discussed.

Task of Peer Assessment 5: Resistance of the wire. In this task, students were instructed to define the resistance of a wire and define how its length affected the current in a circuit. The teacher provided the necessary equipment for the task. This task was similar to the one that measured the speed of the rover, as everyone was supposed to produce an individual lab report of an inquiry conducted in groups.

The peer assessment interventions were embedded in the curriculum, and they lasted three school semesters. Students underwent training prior to the peer assessment, and a formative peer assessments were later implemented in several ways. The interventions were introduced as normal classroom activities. The time frame of interventions and their contributions to sub-studies are illustrated in Figure 5.

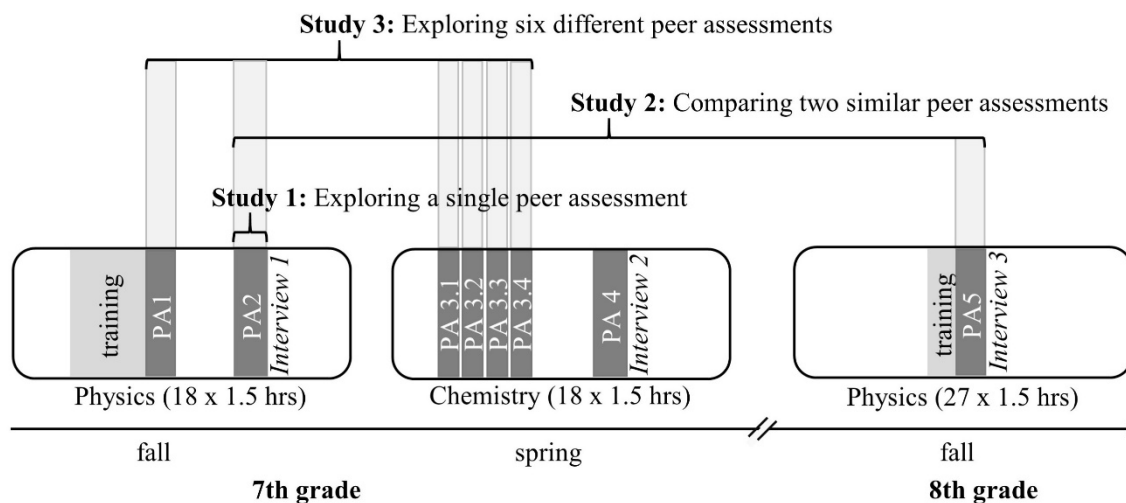


FIGURE 5: Timeline of the training sessions, peer assessments (here abbreviated to “PA”), and interviews

The interventions proceeded without major complications. During the class discussions, students remained focused and seemed to find the topics interesting. The aim of the discussions was not to find a correct answer but to encourage the students to share their thoughts. The training was adaptable, as its intent was to respond to the students’ needs. For example, when peer assessment was practiced for the first time, the students became upset with critical feedback and did not know how to deal with it; the issue was therefore discussed before Peer Assessment 2. During the written tasks and peer assessments, the teacher and I circulated throughout the classroom, discussed with students who needed help, and encouraged them to use and trust their own judgement. Even though the sessions were arranged so that the students were equipped with the essential information, discussions with working groups and individual students appeared influential and necessary.

4.4.1 Peer assessment training

I planned the training by applying prior research on peer assessment. The training included class discussions, written tasks, and one peer assessment. Over six weeks’ time, there were seven 10- to 45-minute sessions and one 30-minute session for the eighth grade students. The training included the elements set out below, and these are described in more detail in this study’s appendices.

Training session 1: Reflection. The goal of the task was to make students reflect on assessment and notice its summative and formative purposes. Students were asked to provide a written answer to the question “What is the use of assessment at school?” The answers were discussed with the whole class and written on the blackboard under the labels “Helps learning” and “Measures learning” or in between.

Training session 2: Reflection. The goal was to understand that feedback is meant to help students and that the quality of feedback affects its efficacy. First, students were asked to provide a written answer to the questions “What kind of assessment helps you?” and “What kind of assessment advances your learning?” Second, they were introduced to the idea that good quality feedback helps people reach their goals, and if it does not, the feedback is useless. Finally, cards with different types of comments were shown to the students, and the usefulness of each comment was discussed. The following ideas were presented with the cards: feedback should be correct; feedback should help the receiver move forward; and feedback’s usefulness is related to one’s skills and goals.

Training session 3: Written task. The goal of the task was to consider what makes good assessment criteria for an inquiry project. The context of the task was an inquiry project that they were about to begin. After the teacher had introduced the project, students were instructed to create assessment criteria on a blank assessment sheet. To assist students in understanding the potentially new concept of criteria, they were provided example criteria from a totally different topic: brushing teeth. In the end, students’ ideas for criteria were discussed and their assessment sheets collected for later use.

Training session 4: Written task. The goal of the task was to further consider the qualities of assessment criteria and practice comparing them to a work. Students were given the criteria they had previously created and a copy of the work of a fictional student, and they were asked to assess that work using their own criteria. After, there was a classroom discussion about how the assessment felt and which criteria worked. In the end, students were asked to count the points they had given to the fictional student and write their totals on the blackboard. The difference in results was discussed.

Training session 5: Self-assessment. The goal of the task was to practice comparing the work to criteria. The context was the previous inquiry project. This time, the teacher provided students assessment criteria, and students assessed their own work according to those criteria. Students were encouraged to discuss their choices in their working groups.

Training session 6: Discussion. As a warm-up for the first actual peer assessment (Peer Assessment 1), the students were reminded of the features of quality feedback with examples of helpful and unhelpful feedback, and these were discussed with the whole class. Right before the beginning, students were reminded about the rules of peer assessment: no laughing at or loud comments about the work. They were reminded that the purpose was for them to help each other move forward and learn to give feedback.

Training session 7: Discussion. The goal of the task was to provide students with new strategies to react to feedback. Before peer assessment, the topic of receiving feedback was discussed with example cards. The idea was to realize that feedback can be bad or rude, and in those cases, ignoring it and moving on is the best thing to do. In addition, three reasons for using peer assessment were presented to motivate students.

Training session 8: Classification of feedback (eighth grade). The goal of the task was to remind students of the qualities of useful feedback. This training took place before the eight-grade peer assessment (Peer Assessment 5). The instructions were adopted from Tasker and Herrenkohl (2016). I had collected students' real feedback from the previous year's peer assessment (Peer Assessment 2). The comments were delivered to the student groups, and they were asked to categorize and label them conveniently.

4.4.2 Peer assessments

Peer assessment was implemented eight times in students' physics and chemistry studies after the first six training sessions (see Figure 5), and each time, it was tied to an inquiry-based activity. The purpose of the assessments was to advance learning, not to measure it, and this was emphasized to students. Formativity was accomplished by giving students the possibility of improving their work after receiving feedback (peer assessments 1, 2, 4, and 5) or using peer feedback with similar subsequent tasks (peer assessments 3.1, 3.2, 3.3, and 3.4). The peer assessment tasks and descriptions are described in Table 3 and in more detail in the appendices.

TABLE 3: Descriptions of the tasks of and instructions for peer assessment (here abbreviated to “PA”)

	Time used	Assessor and assessee arrangement	Assessed task	Assessment criteria	After the PA
PA1	Task: 40 min. Assessment: 30 min.	Groups circulated throughout the classroom and assessed one another’s work.	Group technology project planning and modeling of a rover that moves on its own (Task 1).	Assessment criteria displayed on a whiteboard. Feedback written on Post-it Notes.	Students had an opportunity to modify their plan right after the PA and during the building of the rover.
PA2	Task: 3 hrs. Assessment: 45 min.	Individuals assessed other individuals’ work (pairing planned by teacher and researcher).	Individually written inquiry report on defining the speed of a rover (Task 2).	Provided assessment criteria with three-choice rubric and an opportunity to provide written comments related to each criterion.	Students had an opportunity to revise their report before returning it to the teacher for summative assessment.
PA3s 3.1 3.2 3.3 3.4	Tasks: 4 at 15-30 min. Assessments: 4 at 5 min.	Students in a working pair assessed each other.	Chemistry inquiry conducted in pairs (four different inquiries and PAs; e.g., examining which substances dissolve in water; Task 3).	Assessment criteria with three-choice rubric and a requirement to provide at least one positive feedback comment.	Students marked their agreement with the feedback by circling the most suitable of the four options. The feedback sheets were returned to me or the teacher.
PA4	Task: 15 min. Assessment: 10 min.	Individuals assessed another individual’s work (pairing planned by teacher and researcher).	Individually developed tricky inquiry plan: how to separate salt, sand, iron power, and sawdust (Task 4).	Provided assessment criteria with two choices (“yes” or “no”) and an opportunity to provide written comments related to each criterion.	Students had an opportunity to revise their reports before returning them to the teacher, who checked the plans. The inquiry was later conducted in groups.
PA5	Task: 3 hrs. Assessment: 30 min.	Individuals assessed another individual’s work (pairing planned by teacher and researcher).	Individually written inquiry report: exploring the effect of the length of a wire to the resistance (Task 5).	Provided assessment criteria with three-choice rubric and an opportunity to provide written comments related to each criterion.	Students had an opportunity to revise their reports before returning them to the teacher for summative assessment.

4.5 Data and data collection

The aim of the data collection was to gather a rich data set without interfering in the flow of classroom activities. The data contained information about the progression and details of the peer assessments and peer assessment training. The data included audio recordings of each student pair during the lessons, my field notes, students' original and revised works, written peer feedback, and semi-structured interviews.

I participated in majority of the seventh-grade students' physics and chemistry lessons, and I participated in the lessons that related to peer assessment training or peer assessment with the eighth-grade students. Altogether, I spent approximately 90 hours in the classroom, representing 30 lessons in both classes lasting 90 minutes each. After the students were seated, I set audio recorders on each student pair's table. Students soon became accustomed to the recorders and generally ignored them, though some students played with them, interviewed each other, sent me silly messages, and, probably accidentally, turned them off. After distributing the recorders, I made field notes about the students' attendance and seating order and continued taking notes on the progression and events of the lesson. I also made notes about the students' engagement in learning and wrote down questions about and reflections on the flow of the intervention.

I scanned students' written work and feedback sheets related to peer assessment and photographed the products of inquiry projects. When students had a chance to improve their work after the peer assessment, I scanned both the original and revised work. I tried to keep data gathering as unobtrusive as possible to support the naturalness of the study.

I interviewed the students individually soon after peer assessments 2, 4, and 5 and recorded the interviews. The students who had no previous exposure to peer assessment were left out because the conversation concerned peer assessment. Before the interviews, I read the students' work and feedback and marked the revisions they had made. In semi-structured, stimulated recall interviews (Ryan & Gass, 2012), I showed the others the copy of the work they had assessed, the feedback they had provided, the feedback they had received, and their original and revised work, and I used these documents as a basis for the conversations. As an icebreaker, I asked them about their thoughts of making the original inquiry-related work. Then I asked about their opinion of the work they had assessed and the feedback they had provided for others. I moved on to the feedback that the students had received and ended on the changes, if any, to their work and the reasons for making or not making those changes. If the students brought up other themes, we discussed them as well, and this often changed the order of the questions. Most of the students shared their thoughts openly, but some needed encouragement. The fact that I was familiar to the students helped in carrying out the conversations. I was conscious of the danger of leading their thinking while encouraging them to talk, and I tried to avoid it;

using students' own documents in the interview helped in leading the conversation neutrally to their experiences.

From my perspective, authenticity was achieved well in this study and the results paint a good picture of the reported aspects of peer assessment. I consider the findings valid and informative, but at the same time, they must be interpreted in their context. In addition, the findings are not the whole truth about peer assessment. Different methods might have offered other findings, or a different researcher might have approached the same data from a different perspective, uncovering alternative sides of peer assessment.

4.6 Analysis

Analyzing case studies is difficult (Rowley, 2002) because there are no clear guidelines on how to do it. The analysis was inductive and driven by the data, although the first part of Study 2 was driven by theory. In Study 1, I analyzed the data inductively using a method I developed for this particular purpose called "pathway analysis." In Study 2 and 3, I followed Braun and Clarke's (2006) approach to thematic analysis.

I transcribed the participants' interviews and group discussions during peer assessments and compiled each student's data into a set that contained the discussions, the interviews they had participated in, the artefacts they had produced and assessed, and feedback they had received and provided. Thus, each data set contained broad information about the student's journey through the intervention.

In Studies 2 and 3, I continued the analysis by exploring the data and attaching descriptive labels to essential data chunks (Miles & Huberman, 1994). The data came in several forms, and in many cases, labelling required interpreting extracts alongside other documents. I created the categories through an iterative process that required multiple rounds of labelling. I then examined and adjusted the labels before again coding the data with new labels. When the labels appeared to form categories, I went through the data extracts of each category to examine their coherence, and I wrote descriptions of the categories. The process continued until recoding no longer brought about essential differences to the categories or their descriptions (Braun & Clarke, 2006). To avoid biased interpretations, I discussed the findings with other researchers (Yin, 2009). As a final step, I named the categories. I used Atlas.ti for coding and data management. The program was useful for transcribing the audio files, creating the descriptive list of codes, and modifying the list during coding. Additionally, it allowed me to retrieve all data extracts for each label for examination.

Each study's analysis had its own characteristics. As I was broadly interested in peer assessment, the analytical methods were not fixed. All analyses required a certain level creativity and a tolerance for trial and error. The characteristics of each study are described below.

Study 1. The analysis of students' pathways through peer assessment demanded eliciting information from different sequences of the peer assessment process: working with the original task, providing peer feedback, receiving feedback, revising work, and experiencing other benefits. The information from different sequences came from different data forms, and therefore, the analysis required different levels of interpretation. After creating the sequences and their values through inductive coding, I constructed a pathway for each student through all the sequences (Figure 6).

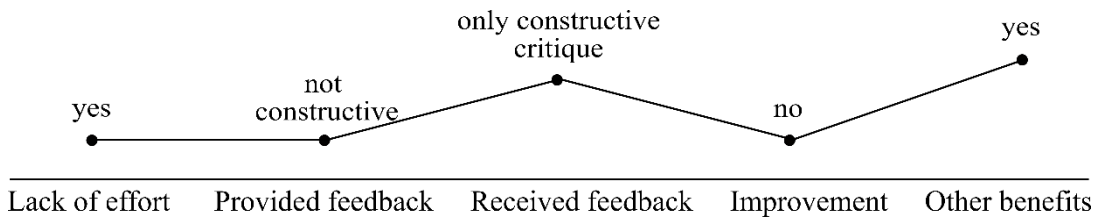


FIGURE 6: An example of students' pathway through peer assessment. This student lacked effort in the original work, he did not provide constructive feedback, he received only constructive critique about his work, he did not improve his work, and he experienced other benefits than improving work.

The reliability of coding was tested with peer coding. Four of five stages of peer assessment were peer coded, and the agreement between me and two other researchers was approximately 80%, which appeared good in qualitative coding that required a high level of interpretation. Moreover, the differences in coding were negotiated. One sequence was based on criteria-based observations of students' efforts in a specific task and could not therefore be peer coded. However, research shows that observations are relevant, as teachers' ratings of students' efforts correlate positively with students' own reports (Zhu & Urhahne, 2014). After coding, I grouped the students' pathways according to the benefits (both improvement and other benefits) of peer assessment that the students experienced. I compared individual students' experiences from different groups to examine and describe the factors that influenced the benefits of peer assessment.

Study 2. The analysis was comprised of three parts. The aim was to recognize the features of feedback literacy that appeared in the data and to adjust the features of Carless and Boud's (2018) framework to the context of formative peer assessment. In a theory-driven analysis, I discerned three categories of feedback literacy skills that students showed during peer assessment. In the second part of the analysis, the aim was to examine students' skills more closely. During the first part of the analysis, I noticed that students had varying feedback literacy skills, and I defined and described them via an iterative process. The analysis was driven by data, but its scope was limited to the previously identified categories. Within each category, I formed case groups with similar skill levels; described them; and with a sensitivity to the theory, organized the groups from

the most basic to the most advanced, creating a criteria-based rubric for feedback literacy skills. A simplified example of criteria is shown in Table 4.

In the third phase of the analysis, I used the criteria that were used to evaluate students' feedback literacy in the midway of the seventh and eighth grades, and I thus defined the development of students' feedback literacy during a year. I then engaged in peer-debriefing (Onwuegbuzie & Leech, 2007) with two other researchers to test the levels and categories. I prepared the full data set of five students for discussion with the other two researchers, who explained their views on my coding, and we carefully discussed the discrepancies.

TABLE 4 A simplified criteria-based rubric for one category of feedback literacy. If a student, for example, did not make any changes to their work on seventh grade after peer assessment, but on eight grade made light changes, they moved from level 1 to level 2 in this category of feedback literacy.

<u>Engagement in making revisions</u>		
<u>Level 1</u>	<u>Level 2</u>	<u>Level 3</u>
No interest in feedback	Reading feedback	Active interpretation of feedback

Study 3. This analysis was the most straightforward and followed the procedure described in the beginning of this subsection. I examined the data, looking for and labeling extracts that related to forms of students' agency during peer assessment. I used Gresalfi et al.'s (2009) definition of agency to identify the extracts: "An individual's agency refers to the way in which he or she acts, or refrains from acting, and the way in which her or his action contributes to the joint action of the group in which he or she is participating" (p. 53). Again, during the analysis, peer debriefing with the two researchers was used to test and discuss the categories. The categories' relationships were elaborated with a thematic map (see Braun & Clarke, 2006) and the identified forms of agency were related to the positions of assessor, assessee, and group member. Therefore, in the last phase, we examined and compared the forms of agency in each of position.

5 OVERVIEW OF THE ORIGINAL STUDIES

5.1 Study 1: Pathways through peer assessment: Implementing peer assessment in a lower secondary physics classroom

Study 1 aimed to explore students' pathways through peer assessment. What kind of pathways did students follow, and which factors worked for and against productive peer assessment? In this study, I systematically observed both roles of peer assessment. Altogether, 22 seventh grade students joined all sequences of peer assessment and were interviewed afterward. Five sequences of peer assessment were explored: working with the original task, providing peer feedback, receiving feedback, revising work, and experiencing other benefits, such as learning about one's strengths or learning something for the future. In inductive analysis, I created categories for students' experiences in each of the five sequences, and after, I coded their experiences in each sequence based on those categories. I presented the pathways through the sequences as graphs and divided different types of pathways into groups. Finally, I investigated exemplary student cases from each group, compared them to each other, and described the cases and their differences.

Students had four different kinds of pathways through peer assessment (see Figure 7). In Group 1, students improved their work after peer assessment and experienced other benefits. In Group 2, students did not improve their work but experienced other benefits. A distinct subgroup of Group 2 were students that did not originally put effort into the task. Their pathways formed Group 3. Group 4 was comprised of students that did not improve their work after peer assessment or experience any other benefits.

The comparison of students' experiences between the groups revealed three factors that worked for beneficial peer assessment: 1) putting effort into the original task, 2) receiving constructive criticism, and 3) understanding the formative nature of peer assessment. Both improving one's work after peer

assessment *and* experiencing other benefits (Group 1) required the existence of all three abovementioned factors. Nevertheless, not all factors were needed for students to either improve their work *or* experience other benefits.

One might think that because of the complexity of peer assessment it is up to chance if an individual student benefits from peer assessment. However, Study 1's findings show that benefitting from peer assessment depends only somewhat on other participants and that students' own efforts and understanding play a significant role in enabling and promoting productive peer assessment. The findings affirmed that peer assessment requires an interplay between individuals and the environment, and to explain peer assessment's outcomes, both aspects must be considered.

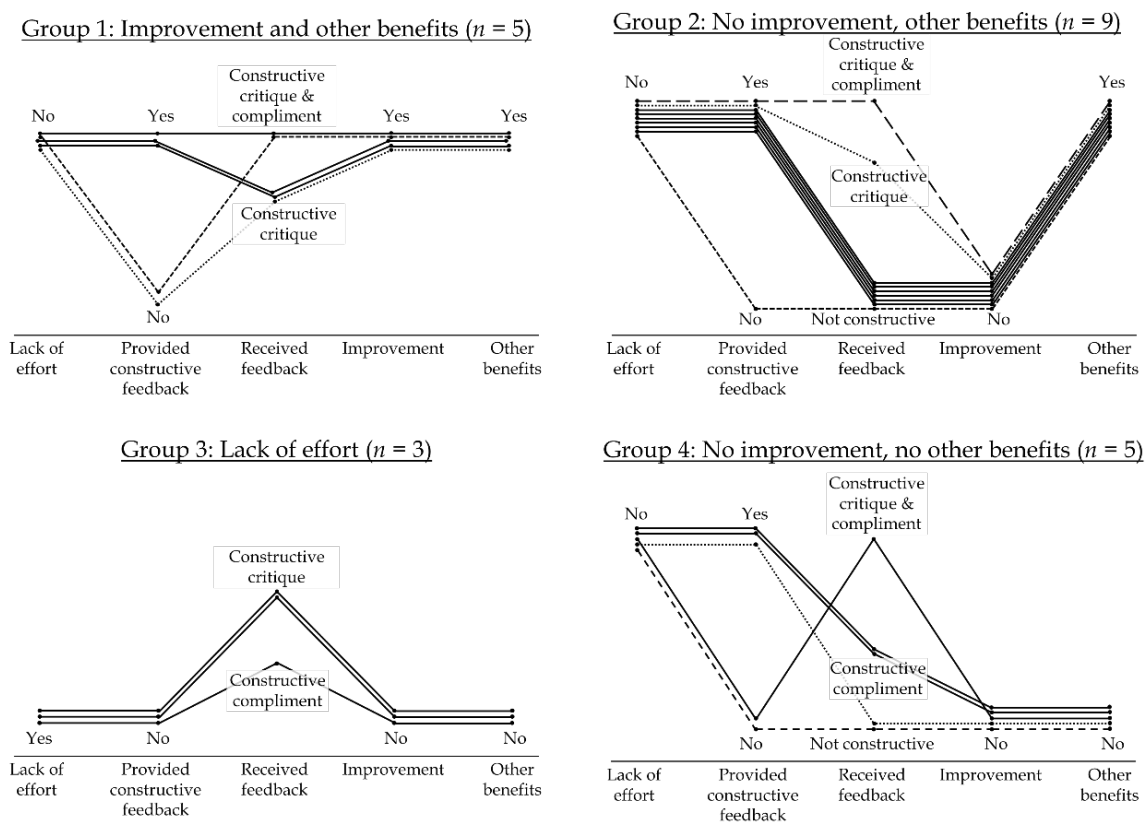


FIGURE 7: Students' pathways through peer assessment

The findings revealed the significance of students' own goals. If students did not put effort into the original task, the feedback was ineffective. It could not close the gap between students' current understanding and their goals if they did not have any. Therefore, discussions about the objectives are necessary in classrooms. Without goals, the feedback on performance is pointless.

Critical feedback turned out to be another important issue. Constructive criticism appeared effective and, in this study, necessary for improving work after formative peer assessment. However, providing, receiving, and valuing criticism was challenging for students, especially if they did not understand the formative purpose of peer assessment. It was crucial that students understood

that formative assessment is meant to assist in learning, not to judge or criticize them. Therefore, as a part of peer assessment training, students should learn to provide constructive criticism and to appreciate its value in learning.

5.2 Study 2: The development of secondary students' feedback literacy: Formative peer assessment as an intervention

Study 2 aimed to investigate students' feedback literacy skills and the development of those skills (Carless & Boud, 2018). The results of Study 1 had shown that students struggled with certain features of feedback literacy. In this study, I investigated the scope of students' feedback literacy skills and examined how they had developed after a year of practicing peer assessment in physics and chemistry.

To learn about students' feedback literacy, I organized two similar implementations of peer assessment, the first in the midway of the seventh grade and the second a year later (PA2 and PA5 in Figure 5). Altogether 13 students had participated in all parts of these two peer assessments and in most of the other peer assessments as well. I investigated these students' feedback literacy from interviews that were interpreted alongside other data sources, including students' written work, their received and provided feedback, the work they had assessed, field notes, and audio recordings of the lessons. Using a thematic analysis, I identified three categories of feedback literacy and created criteria for three skill levels in each category. Using the criteria, I coded students' skills in their seventh and eighth grades and examined their development.

The skill levels described the scope of lower secondary students' feedback literacy in each category: conception of the purpose of feedback; engagement in and interpretation of received feedback; and activity in making revisions (see Table 5). *Conception of the purpose of feedback* referred to students' expectations of feedback. Students on the first skill level were not interested in any kind of feedback. Students on the second level held a summative view of feedback and hoped it would tell them that they had succeeded. Students on the third level understood that peer assessment was supposed to advance their learning and welcomed critical feedback. *Engagement and interpretation of received feedback* referred to students' attitudes and activity in receiving feedback. Students on the first level were not interested in feedback and ignored it. Students on the second level read the feedback but did not actively interpret it, which led to, among other things, students abandoning feedback when they did not understand it or accepting feedback even when it seemed wrong. Students on the third level actively interpreted or evaluated the feedback. *Activity in making revisions* referred to how students revised their work after the peer assessment. Students on the first level did not revise their work. Students on the second level revised their work but only by adopting the changes proposed in the feedback. Students on the third level were active in revising their work and did not expect the

feedback to fully guide them through their revisions; these students practiced new skills, learned new knowledge, or sought more feedback.

The levels of feedback literacy revealed that not all students were prepared to use peer feedback for learning. Some students were not interested in feedback, were offended by criticism or did not know how to proceed with feedback. In such cases, improving the quality of feedback or increasing its quantity would not have solved the issue because the students needed a better understanding of the purpose of feedback and the process of receiving it.

TABLE 5 The criteria for the categories of feedback literacy

Category	Level 1	Level 2	Level 3
Understanding of the purpose of feedback.	No appreciation of feedback. Student expressed that they did not appreciate feedback.	Appreciation of positive feedback. Student expressed that receiving positive feedback was satisfying and desired.	Appreciation of corrective feedback. Student expresses that receiving corrective feedback was desired.
Engagement and interpretation of received feedback.	No interest in feedback. Student expressed that they were not interested in engaging the feedback.	Reading feedback. Student expressed that they had read the feedback, but they did not show signals of actively interpreting it and comparing it to their work, or they had problems with selectively rejecting the feedback.	Active interpretation of feedback. Student actively interpreted the feedback by comparing it to their work and, if they deemed it necessary, rejecting it.
Engagement in making revisions.	No improvement. Student did not improve their work.	Superficial improvement. Student made improvements that did not require much effort or the learning of new knowledge or skills.	Investment in improvement. Student improved their work as a result of their own effort by, for example, seeking new feedback, guidance, or knowledge or by practicing new skills.

The examination of students' skills in the seventh and eighth grades revealed that their feedback literacy developed over the year. Ten of the 13 students developed in one or more of the three categories of feedback literacy, although one student dropped a level, and five students developed in two or three categories. A good example case of development is a student who received only positive feedback in both the seventh and eighth grades. In seventh grade, she was content with the feedback and said that it was encouraging and peer assessment was fun. In a similar situation in the eighth grade, she said that she had expected feedback that showed her how to improve her work. It appeared that, over the year, she had

learned to understand feedback's value for development. Another example is a student who did not know what to do with critical feedback in the seventh grade. He neither used it to improve his work nor consciously rejected it. In the interview, he explained that he had taken the feedback as guidance for the future, which contradicted the content of the comment because it related to the specific work. Then, in eighth grade, the student evaluated the feedback he received and selectively rejected some comments and used others to develop his work. Over the year, he had learned how to engage with and actively interpret received feedback.

A noteworthy point about the development of students' feedback literacy is that they were not novices when their skills were determined in seventh grade: They had already trained in peer assessment skills and practiced peer assessment once. Based on this point, it can be said that the development of feedback literacy takes time. Consequently, because feedback literacy promotes productive peer assessment, making peer assessment a functional learning tool may require patience. Even if peer assessment appears unproductive at first, it can be of service in terms of developing students' feedback literacy. Consciousness of the prolonged development may comfort teachers who are experimenting with peer assessment in their classrooms.

The observed development of students' feedback literacy suggests that peer assessment not only helps students learn subject knowledge and skills but also advances their feedback literacy. Therefore, peer assessment is one way of augmenting students' understanding of feedback processes. Most peer assessment research at the secondary level has concerned subject skills, but this research demonstrates that other approaches are also worthwhile. Feedback literacy is an important life skill and intrinsically valuable, but its development may also indirectly support students' learning. In addition, the results demonstrate that lower secondary students can already develop feedback literacy. Thus, students should be introduced to feedback literacy long before higher education, for example, via formative peer assessment.

5.3 Study 3: How do lower-secondary students exercise agency during formative peer assessment?

In Study 3, I examined the concept of agency to understand the sociocultural dimension of peer assessment. Peer assessment is a social activity, and its purpose is to use interactions among students—spoken or written—to promote learning. The aim of Study 3 was to explore how students exercise agency during peer assessment.

The data came from the first six peer assessments (peer assessment 1, 2, 3.1, 3.2, 3.3, and 3.4) implemented during the seventh grade physics and chemistry studies. I used the most authentic data—students' work, classroom discussions, and written feedback— as primary sources, and I turned to interviews and

observations to complement and explain the findings. Seven students participated in all interventions, and I chose five of them for further analysis. These five students' data were of higher quality than the other two, as they were more inclined to speak to their friends and teacher during the lessons, whereas the two students that were left out did not add to the variety of the data.

Using a thematic analysis, I identified and categorized data extracts containing information on students' agency, and described and named the categories. With thematic map, I elaborated the categories' relationships. As a result of analysis, I identified 12 forms of agency as associated with three positions: group member, assessor, and assessee (see Figure 8).

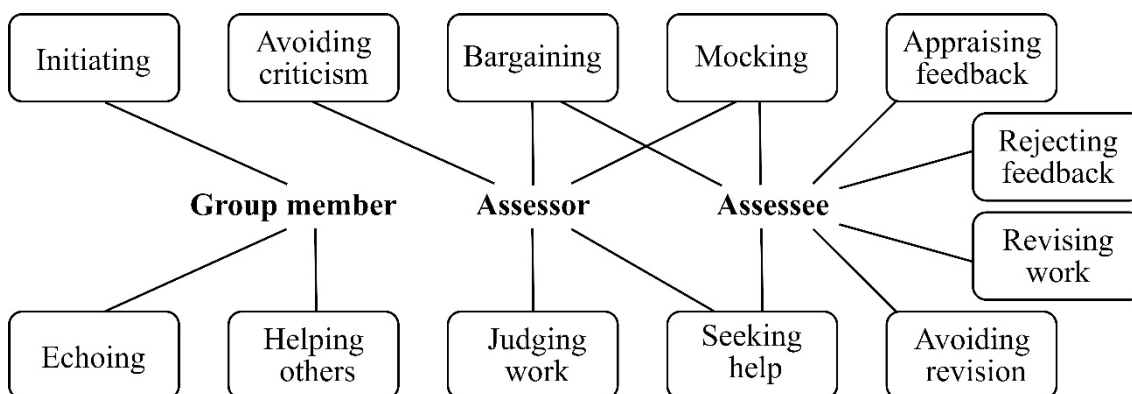


FIGURE 8: The forms of agency and the positions in which they were exercised

The arrangements of peer assessments influenced students' positions in relation to other students and the agency they exercised. When students acted as assessors and assessees in a group, they could practice the agencies of initiating and echoing. Then, students who experienced difficulties with providing or receiving feedback could receive subtle support from their groupmates by following their initiatives. For example, they slightly elaborated others' feedback comments or used identical ones later with other pieces of work. In individual peer assessments, students needed to be responsible for their progress. If they struggled with the task, they needed to ask for help or find another way of coping, which often meant refraining from doing the task by exercising some other form of agency. For example, students avoided judging others' work by providing only positive feedback, and they escaped revisions by focusing on other activities. When students assessed each other in pairs, they could discuss the assessment during the feedback provision in the assessor and assessee roles, which assisted in the task. As a downside, the less formal arrangement led some students to bargain about the feedback, and thus, avoid the formative aim of peer assessment.

Closer examination and comparison of students' agency showed that they were unequally challenged by peer assessment. Throughout the year, two of the five students—Rachel and Lucas—did not seem to have difficulties with taking initiative or appraising others' work or feedback; two other students—Maggie and Nathan—needed help to exercise such agencies. When peer assessment was conducted in groups, Maggie and Nathan did not express original ideas; instead,

they echoed and elaborated those that others had initiated. As individual assessors, they had similar difficulties but different ways of dealing with them. Maggie was more persistent as an assessor, Nathan as an assessee. Throughout the year, Nathan refrained from criticizing or advising others and provided only positive feedback. In contrast, as assessor, Maggie wanted to judge the work properly and give critical feedback when appropriate, but she could not do this without help. She persistently asked for support from the teacher and students, and thus, gained agency to provide valid feedback. Returning to Nathan, as assessee, he sought help with appraising the feedback, and this enabled him to improve his work. In contrast, Maggie avoided processing the critical feedback, did not ask for help, and did not revise her work. The fifth student, Nick, differed from the other four, as he did not have difficulties with expressing his ideas or appraising others, but as assessor and assessee, he exercised mostly maladaptive agency, meaning that his actions did not advance his or other students' learning.

According to the results, acting as assessor and assessee fit better with the roles of some students than others. Following from this, if students appear incapable or unwilling to help their peers or accept help during peer assessment, it does not necessarily signify that they have a poor attitude or lack skills. Rather, one possible explanation is that they struggle with exercising agency. Therefore, to ensure productive implementation of peer assessment, students also need support with their agency.

6 DISCUSSION

The findings of this thesis have implications for theory and practice, and these are discussed in the next two subsections, followed by considerations about the study's trustworthiness, ethical issues and limitations, and suggestions for future research.

6.1 Owning the feedback process and practicing feedback skills

In this study, I explored students' approaches to peer assessment and feedback. In Study 1, an inductive examination revealed that the productivity of a single peer assessment is contingent on students' feedback literacy, particularly their ability to judge their peers' work, produce constructive criticism, and understand and appreciate formative assessment. An investigation into the features of feedback literacy in Study 2 showed that seventh graders are already able to develop feedback literacy. Feedback literacy's value has been recognized in universities (O'Donovan, 2017) and workplaces (Carless & Boud, 2018). While most people will not attend higher education, feedback literacy skills are relevant to everyone, and students' feedback literacy should be fostered and further researched in secondary education.

Even more importantly, this study was the first to identify and describe criteria for students' feedback literacy. Prior studies on the development of feedback literacy did not use explicit criteria but described changes in individual students' skills and attitudes (Han & Xu, 2019b; Hey-Cunningham et al., 2020). With criteria, the features of feedback literacy become concrete and applicable to researchers, students, and teachers. For researchers, the criteria enable an understanding of the phases of development and allow for tracking and comparing the development across disciplines and school sectors. The criteria are also useful for students' self-reflection and self-assessment because they make students aware of their own attitudes and skills, and they communicate how students could proceed (Andrade & Du, 2007). Teachers could show the criteria

to students and discuss them before delivering feedback. With the criteria, they could point out, for example, that embracing criticism is difficult but desirable or that merely reading feedback is not the most effective way of using it. For educators, the criteria offer an understanding of the range of students' skills and thus assist in organizing peer assessment training, identifying students' strengths and needs, and adjusting support.

The findings of Study 2 showed that students are not necessarily inclined to read feedback or use it for improvement. According to Study 1, students who are not interested in a task ignore even useful feedback about it. As the purpose of feedback is to help students reduce the gap between their current and desired understandings (Hattie & Timperley, 2007), feedback has little effect if students have no goals or are not interested in pursuing them. In such cases, there is no point in pushing feedback; rather, teachers could help students to find a goal that is worth pursuing. Likewise, critical or guiding feedback cannot help students achieve their goals if they consider it offensive. Teachers could help students with this issue by communicating the message that critical feedback is a natural part of learning and does not signify failure. Distinguishing the different purposes of formative and summative feedback would be helpful for students. As noted in Study 3, some students focused on their marks during peer assessment even when marks were not given. This tendency is probably the result of the prevalently summative assessment practices in Finland (Atjonen et al., 2019). Although the idea of formative assessment was introduced to students with peer assessment, they were unaccustomed to it, and apparently, they could not digest it immediately.

The results showed that students are rather passive in the feedback process. In this study, even when they were interested in feedback, most students did not invest in interpreting it; instead, the students expected that feedback would tell them unambiguously how to proceed. This mindset likely comes from students being accustomed to receiving teacher feedback, which is not generally questioned or rejected (O'Donovan, 2017). Assessment has traditionally been the purview of teachers (Boud, 2014), with students as the objects of assessment (Boud & Falchikov, 2006). If students are not accustomed to judging the feedback they receive, they should be encouraged to do so. Teachers could discuss different ways of reacting to feedback, such as interpreting its meaning, using it immediately, writing it down for the future, or rejecting it. If students are guided to choose whether and how to act on feedback, they may begin to see critical feedback as a gift that can advance learning instead of a burden that always signifies more work.

It is unclear to what extent the use of peer assessment affects the development of students' feedback literacy. However, peer assessment offers opportunities to have discussions about the features of feedback literacy, which is one potential reason for the development of students' feedback literacy. When comparing this and other studies on the development of feedback literacy (Han and Xu, 2019b; Hey-Cunningham et al., 2020), the common elements are explicit discussions about feedback processes and the use of peer feedback. Hence, this

study supports earlier findings about the training for and implementation of peer assessment as a tool to promote the development of students' feedback literacy.

Contextual elements influence students' feedback literacy (Chong, 2020), particularly assessment culture. Peer assessment and students' feedback literacy are not practiced or developed in a vacuum, but schools' other assessment practices mold students' understanding of assessment and feedback. In Finland, many elements of assessment maintain students' passive role. On average, students are more aware of summative than formative assessment and teachers' feedback comes most often in the forms of grades or points (Atjonen et al., 2019). Commonly used technology-enhanced feedback is important to students, but it mainly takes the form of teachers' one-directional messaging, and it guides students to focus on a superficial level of behavior (Oinas, 2020). Even student-teacher-parent discussions, which are aimed to be participative, often turn out to be teacher-centered provision of information and neglect students' active participation (Luukkonen, 2020). It is contradictory if students are given full responsibility to assess other students' work in one situation but not expected to comment even on their own assessment in another. Therefore, it would be useful not just to add participative assessment elements but also to examine all the pieces together. Better alignment between schools' feedback practices would support the development of students' feedback literacy.

Teacher-centered summative assessment puts students in a passive role and may therefore hinder the development of their feedback literacy. When both formative and summative practices are used simultaneously, summatively assessed tasks tend to attract students' attention (Silseth & Gilje, 2019). Formative assessment's potential of disrupting the traditional teacher-centered learning culture has been questioned (Nieminen, 2020), which is relevant to this study because though peer assessment was used formatively, the students received summative grades at the end of the learning modules. In self-assessment, the contradictions between formative and summative practices can be tackled using summative self-assessment—that is, letting students decide their own grades (Nieminen & Tuohilampi, 2020). However, with peer assessment, formative practices seem to provide a deeper orientation to learning than summative practices, as the latter focuses on grades instead of learning (Panadero, 2016). The results of this thesis show that formativity does not distort peer assessment or hinder the development of students' feedback literacy. The alignment of formative peer assessment with summative assessment—for example, by letting students improve their work after peer assessment but before the teacher summatively evaluates it—may even accentuate the focus on learning and development of students' work. This is not to say that this study's assessment practices and its balance of formative and summative assessment were optimal. An interesting topic for future research would be a comparison of peer assessment interventions and the development of students' feedback literacy in different assessment cultures.

Peer assessment is a concrete way of sharing the responsibility for learning and assessment with students and molding the culture of the classroom into

something more participatory and student centered. Even though peer assessment might appear awkward or technical at the outset, it can gradually lead students toward relational agency (i.e., the ability to seek and use others' help and support others in mutual attempts to learn; Edwards & D'Arcy, 2004).

6.2 Taking on new constructive roles

Although peer assessment is a social activity, its social side has received surprisingly little attention from researchers (Panadero, 2016). I explored peer assessment from a sociocultural perspective and found that students' agency influences their participation. Even though all students were placed in the same classroom to do the same peer assessment tasks, they experienced unequal challenges that did not derive from cognitive issues or their attitudes but from their agency. This derives from agency's construction via experience (Emirbayer & Mische, 1998), where students are expected to act and engage in classroom activities in a similar manner to what they have previously done (Gresalfi et al., 2009). The students who are typically initiative-taking or critical, or who usually help others, can act in their normal roles as assessors and assessees, and their peers will probably accept their critique and guidance. However, those students who are less initiative-taking need to act against their accustomed role, which is challenging and potentially less approved of by other students.

The finding related to the agency requirement during peer assessment is vital. It informs about the influence of individual differences on peer assessment; according to Panadero (2016), this is an entirely unexplored territory. It also highlights the need for individual support during peer assessment because students are requested to exercise forms of agency that are new to them, and hence, exit their comfort zones – some further than others. Awareness of agency allows teachers to notice the need for support and provide it appropriately by listening and endorsing students' ideas and allowing and encouraging students' discussions with friends. Thus, the students can lean their opinion on someone else's ideas and borrow their agency. When it comes to teachers, their role in facilitating new participatory learning practices is crucial but delicate (Kumpulainen & Rajala, 2017). Without an awareness of students' need for agency during peer assessment, teachers may misconstrue that students' underperformance or misbehavior stems from a negative attitude or insufficient subject skills. Instead of providing students with firm support to act as assessors and assessees, the teachers should encourage and endorse students' thoughts; otherwise, they can undermine students' agency by suggesting that the students are unable to perform peer assessment independently.

From a sociocultural perspective and in consideration of the abovementioned findings, instructing students to perform peer assessment individually and without support is unethical and creates unequal outcomes. Since agency is built on experience, students with lower subject skills are more likely to lack agency as well. Hence, some face a double challenge, as assessing is

cognitively challenging and they lack the agency to do it. This study's findings reveal that without support, struggling students are likely to exercise a less constructive form of agency, as they find it more comfortable. Consequently, these students do not benefit equally from peer assessment. The equity of summative assessment has been widely questioned and investigated (e.g., Brunello & Checchi 2007; Darling-Hammond, 1991; Hilden et al., 2016; Ouakrim-Soivio, 2013; Ouakrim-Soivio et al., 2017). As formative assessment promotes learning (Black & Wiliam, 1998a, 1998b) which is transmitted to summative results, it should be examined through the lens of equality as well. As this study has demonstrated, sociocultural approaches are valuable in exploring the equity. They lead to critical examinations of the dominant ways of implementing assessment and improve understandings of how assessments treats different student groups (Elwood & Murphy, 2015).

Despite its potential pitfalls, peer assessment can advance democracy in the classroom between teachers and students (Gielen et al., 2010) and among students by sharing with everyone the responsibility for and opportunity to help others. It is natural that new practices create tensions between students' accustomed role and the ways they are expected to participate (Kumpulainen & Rajala, 2017). Changes in students' positions, tasks, and relations are gradual and occur as students participate in the classroom practices (Rainio, 2008). As individuals' actions create expectations for students' participation, the trials of new forms of agency during peer assessment can become the new normal and reform classrooms' fixed roles. Peer assessment with appropriate support from teacher and students provides a platform to exercise agency and establish a more democratic learning environment.

6.3 Considerations of trustworthiness and ethical issues

As explained in this study's introduction, my research was motivated by my personal interest and creativity. In qualitative studies, researchers function as research instruments, and to avoid researcher bias, it is essential to reflect own values and experiences of the researched phenomenon (Cope, 2014). My personal commitment even increased the need to become aware of my values, experiences and prejudices. Since researchers' biases, insights, and understandings can be of service, the researcher should not aim for complete neutrality (Anderson & Shattuck, 2012). In thematic analyses, two researchers are not expected to arrive at the same themes (Braun et al., 2016), as their experiences and knowledge shape their understandings and interpretations of the data. However, there are methods to assess and advance the truth value of qualitative research (Cope, 2014; Noble & Smith, 2015; Onwuegbuzie & Leech, 2007), which I will now discuss.

From the outset, I aimed to achieve the basic requirements of any study: commitment, skepticism, and detachment (Norris, 1997). I achieved commitment through prolonged engagement, which entailed understanding the culture, building trust with participants, and conducting the research for long enough to

obtain sufficient data (Cope, 2014; Onwuegbuzie & Leech, 2007). As I had worked for several years at the school, I was familiar with its culture, which also facilitated my building relationships with the students. During the data gathering, I spent over 100 hours (approximately 30 lessons of 1.5 hrs with both classes, plus the interviews) at the school, which gave me a satisfactory understanding of the case. The collection of rich data diminished confirmation bias (Onwuegbuzie & Leech, 2007), and the multiple data sources increased the credibility and truth value of the research (Cope 2014), and allowed me to triangulate the data and weigh the evidence. The distinct data forms supported each other, but in some cases, the data were contradictory, such as when a student said they liked peer assessment in an interview but sounded very stressed and complained about having to assess their peer on a recording. In such cases, I gave more weight to the firsthand data (Onwuegbuzie & Leech, 2007).

While it can be a strength, a researcher's familiarity is also a potential source of bias. One source of bias is researchers' inability to recognize and separate their presumptions (Onwuegbuzie & Leech, 2007). Hence, I took time to reflect on my prejudices and assumptions. Regarding the research topic, I was curious but had no specific expectations. Based on the literature, I assumed that peer assessment might be beneficial for learning, but I found myself equally open to opposing findings. I considered my experience as a teacher as a potential source of biased interpretations because a sense of experience could lead to interpret things in a routinized way. I dealt with the issue by reflecting on my conceptions alone and in an ethics study module and by reminding myself that, according to the research, teachers are only aware of a portion of what happens in their classrooms (Nuthall, 1995). In my own judgement, I succeeded in approaching the classroom and data with curiosity and an open mind. I engaged in discussions with other researchers to reflect my ideas and interpretations (Noble & Smith, 2015). I discussed the research design, the interventions, and data gathering with a group of fellow researchers and reflected on their feedback. Peer debriefings and peer coding with two other researchers during the analysis yielded beneficial criticism and helped me see the study from an outsider's perspective.

Researchers can also bias their results by affecting participants (Onwuegbuzie & Leech, 2007) or letting participants affect the results. I used a number of methods (Onwuegbuzie & Leech, 2007) to reduce potential bias. The study was naturalistic, and the data were collected using the most unobtrusive methods possible. The recorders were used in every lesson, not just significant ones, to help students get accustomed to them, and I copied students' work and feedback after the lessons. However, I was open about my intentions and role, and the students accepted my presence in the classroom. As mentioned before, I avoided taking responsibility for classroom management, and I refrained from reacting to negative behavior and from praising students for accomplishments. Hence, I got more sincere reactions from students, as they did not receive extra attention by behaving poorly and could not disappoint me with low achievements. In the interviews, I was careful not to lead students' responses

with my questions or reactions. In hindsight, a slightly more intrusive style might have been beneficial for eliciting information.

Participants' influence on the results was controlled by inviting the students of two standard physics classes to take part in the research. The students had been assigned to their physics classes using the alphabetical order of their last names as a primary factor. As only two of 31 students refused to join the study, the sample was rather randomized. Inviting two classes to join the study permitted a comparison between the groups and gave me a perspective on the uniqueness of the phenomena I noticed.

The ethical issues also affect the integrity of the study (Connelly, 2016). For one thing, I had to consider the official rules about conducting a research. While planning the research, I followed the guidelines of the Finnish Advisory Board of Research Integrity (2012) and National Advisory Board on Research Ethics (2009). In addition, I used Stutchbury and Fox's (2009) framework for reflecting on ethical issues.

The data were pseudonymized and stored according to the university's guidelines. While reporting the results, only excerpts of students' discussions were published to maintain their anonymity. The findings were published only in English, which required translating what students said and helped in concealing their identity. Permission for the research was granted by the city's educational administration. The students' parents signed a consent form, and students and parents were informed that the decision to participate would not influence students' treatment in the classroom and that they could withdraw their consent at any point with no consequences.

According to Stutchbury and Fox (2009), every researcher must consider their study's potential benefits and drawbacks for participants, organizations, and society. The authors present a list of questions for reflection on the ethical aspects of research, including some that were central to this study, which I discuss below. Since the study was naturalistic and I was familiar with the school, I was able to plan the research so that it did not expose students to abnormalities. Since implementing peer assessment is a mandatory part of the curriculum, the students would have practiced it in any case. Given that I had the opportunity to invest time in planning the implementations, the teacher and students were likely to benefit from the arrangement. The teacher wanted the focus of the research to be on the students because she preferred to stay in the background. Therefore, the teacher's voice is not explicit in the research, but her observations and insights influenced the course of the intervention. I often walked around the school with the teacher during her break-time supervision, and we reflected on the students, the group dynamics and the intervention. I think that the discussions were fruitful for both of us. The teacher's commitment to the study and her routine to plan things ahead were valuable. I considered her as a good match for this research because her teaching met good standards, but at the same time, her practices were not especially distinctive. Naturally, all teachers use their personality in their work, but some teachers do so more obviously compared with others. I did not aim to explore a particular case, but instead, I sought to

look at practices that would be realistically achievable for all teachers. I shared peer assessment materials freely, and they were therefore a potential benefit to all the school's science teachers. As the aim of this study was to contribute to an improved understanding of a rather under-researched topic—particularly in Finland but also internationally—I considered the research beneficial to a larger audience as well.

For participating students, an additional benefit was having a second science teacher in the classroom to support their learning. I aimed to avoid causing harm by not using the students' and teacher's time more than necessary, and I prevented awkwardness in classroom interactions and interviews by establishing good relationships with the students. I met the students in the beginning of their lower secondary school, and as the building, teachers, and practices were all new to them, my presence was not particularly striking. Although I was in a researcher's role, I noticed that the most natural way of interacting with students was my accustomed teacher manner, which included casual discussions, bad jokes, and providing support. I developed good relationships by being open about my role and about the research. For example, in the beginning of the eighth grade, I reminded the students about the study by telling them about my conference trip to Italy and my preliminary findings. In a typical "cool" teenager manner, the students did not react strongly to my presence, but there were signals of their conceptions about me and the intervention. When I introduced an element of intervention, I felt that students became alert and participated with a positive attitude. They also appeared interested in the research and asked questions about characteristics of the researcher's job. In addition, I remember one student expressing his disappointment when I told him that I could not write about them using the students' real names. He would have liked to be known by his name.

Naturally, students did not experience peer assessment as only enjoyable. During the last PA3, one student group made comments in a satirical tone on the audio file about how they "again had an opportunity to assess each other." In the interview, one student politely mentioned that he considered peer assessment artificial and useless. However, the resistance was minor. I construed that participating in a research study was a welcomed, although not highly remarkable, way to spice up students' ordinary school days. In terms of their relaxed way of talking to the audio recorders and me in the class (not necessarily in the interviews), I concluded that they were fine with my presence and my intentions to study their learning. I think that my openness was an essential ingredient in gaining students' trust; moreover, being aware of the study allowed them to ignore it and act unreservedly.

6.4 Limitations and future directions

The main limitations of this thesis are intertwined with its strengths. The naturalness of the study was an intentional choice, but it prevented me from

controlling numerous variables during the intervention and data gathering. School days are unpredictable. For example, when the students ran a Cooper's test, they were exhausted in the next lesson, or when someone left their work at home, it could not be assessed by their peer and the possibility of gaining data from that student for that part of the intervention was lost. The main influence was the loss of data when tasks were interrupted or students were absent from school.

The participants came from the same school and were taught by the same teacher. Hence, the results must be interpreted in their context. Even though the researcher's familiarity with the school and field was a strength in organizing the intervention and analyzing the data, it inevitably guided the interpretations. This is not a criticism of the results; rather, it highlights that another researcher with different practical and theoretical backgrounds could come up with other findings. Hence, this research should be seen as more of a beginning of a discussion than a final word.

The student groups in this study were rather small—it is nationally recommended for experimental science classes to be small—and the researcher participated as an extra resource. Therefore, during peer assessments, more support was available for students than in an average classroom. Consequently, the teacher and I were better able to monitor students' needs for support and provide it when appropriate. In contrast, the awareness of the research made us consider the need to provide support carefully because we did not want to contaminate the results by guiding students to act according to our preferences. In my view, both features positively influenced the implementation of peer assessment. Students' agency was not suppressed with unnecessary support, and we could give time to students when they needed it. Especially when peer assessment was first used, students needed encouragement to write down their thoughts. With less support, more students might have exercised some maladaptive form of agency. As a higher student-teacher ratio limits teacher's possibilities to support students individually during peer assessment, it is even more crucial to prepare for peer assessment in advance with appropriate training and discussion.

Although this study did not investigate students' language skills, participation in peer assessment necessarily required verbal and written skills. Based on the analyses, it is not possible to say to what extent these skills influenced students' performances. Nevertheless, there are indications that language skills were not a dominant feature. From the audio file discussions, I noticed multiple cases in which a quantitative mark combined with a brief comment induced a meaningful discussion or revisions. The provision of short comments was supported by rubrics that offered relating vocabulary. Naturally, there were also lengthier, well-articulated feedback comments written in full sentences. During the intervention, I expected that assesseees would experience such comments as alienating or irritating, especially if assessors used a teacher-like tone. However, I did not notice signs of irritation. It is possible that what I

construed as imitation of the teacher's style appeared as good quality feedback to the students.

The results of this thesis reveal two new research directions. The first relates to feedback literacy. Previous research on peer assessment has had a strong focus on psychometric qualities, such as the reliability, validity, and accuracy of peer assessment, even though there is little knowledge of whether accuracy is even essential to students' learning (Panadero, 2016). When the effects of learning have been researched, the researchers have mainly concentrated on subject skills. The results of this thesis suggest that peer assessment promotes useful attitudes and learning-to-learn skills, which should get more attention in the future. The following questions are worthy of further research: What forms and features of peer assessment promote students' feedback literacy? In which ways and to what extent can the findings of previous peer assessment studies be explained by students' feedback literacy? How does students' developed feedback literacy influence their subject learning?

The second research direction involves the sociocultural phenomenon of peer assessment. I agree with researchers who argue that this direction is under-researched (Panadero, 2016; van Gennip et al., 2009). This thesis showed that a sociocultural perspective can reveal new and unexpected views and findings about peer assessment; the influence of students' agency is definitely not the last. Longitudinal research on students' agency would be highly interesting. It would be of interest to examine whether peer assessment encourages students to transcend the classroom's fixed roles and how it does so. Peer assessment has been proposed to advance democracy in the classroom (Gielen et al., 2010) – but does it do so in practice? In this study, some students lacked the ability to exercise certain forms of agency, but would exercise of such agency become more achievable to them with further experience of peer assessment? If so, are classrooms where peer assessment is used regularly more democratic compared with those where peer assessment is not a common practice – so that expressing ideas, helping others, and accepting help are more widespread among all students?

6.5 Final words

When I began this research, I was unsure of the rationale behind peer assessment, and I worried about the lack of guidelines for its implementation. With my journey now coming to its end, it is time to reflect on my understanding of these issues.

Based on previous research on training for peer assessment and the skills it requires, and considering my finding about the importance of students' agency, peer assessment's implementation definitely requires a profound comprehension and guidelines. Without a shared understanding of the criteria, rules, and aims of peer assessment among students, and without sufficient training and support, the practice is unproductive, unequal, and potentially harmful. To avoid

unnecessary missteps, resources and training should be easily accessible to teachers if they are required to implement peer assessment. Teachers' professional development must be taken seriously. When curricula and instructions change, teachers need knowledge and time to reflect on and adopt new procedures.

Based on this study's findings, peer assessment is more profound than sharing information between peers. During peer assessment, students can learn to process and act on feedback and practice operating in new constructive roles. These are essential elements for activating students as owners of their learning. Peer assessment disrupts the widely held perception of teachers having a monopoly on assessment and feedback. During peer assessment, teachers do not transmit feedback. Instead, they activate students and facilitate their own feedback processes. Therefore, peer assessment can be used to accentuate students' responsibility in learning. In the current learning transformation that encourages students' agency in learning, I consider the most important rationale for implementing formative peer assessment to be the establishment of a student-centered learning culture that focuses on learning instead of extrinsic goals and increasingly appreciates and promotes collaboration.

YHTEENVETO

Väitöskirjassani tutkin formatiivisen vertaisarvioinnin käyttöä yläkoulun fysiikan ja kemian oppitunneilla. Tutkimukseni tavoitteena oli selvittää, mitä vertaisarvioinnin aikana tapahtuu ja siten ymmärtää paremmin sen mahdollisuuksia. Laadullisen tapaustutkimukseni kohteena oli kaksi oppilasryhmää, jotka harjoittelivat ja käyttivät vertaisarviointia fysiikan ja kemian opinnoissa vuosiluokilla 7–8. Seurasin oppilaita ($n = 29$) seitsemännen luokan alusta kahdeksannen luokan syyslukukauden loppuun. Yhteistyössä ryhmien opettajan kanssa suunnittelin vertaisarvioinnin taitojen harjoittelun sekä vertaisarvioinnit osaksi oppiaineiden opiskelua. Tutkimusaineistoni sisältää muistiinpanoni ja jokaisen oppitunnin äänitallenteet (noin 90 tuntia), oppilaiden haastattelut, vertaisarvioitua kirjalliset työt sekä kirjalliset vertaispalautteet. Moninaisen aineiston analyysit olivat laadullisia ja suurimmalta osin aineistolähtöisiä, ja käytin niissä ideoimaani ”polkuanalyysia” sekä temaattista analyysia. Ensimmäinen osatutkimus kohdistui yhteen vertaisarviointiin, toinen kahteen samankaltaiseen vuoden välein toteutettuun vertaisarviointiin ja kolmas kuuteen peräkkäiseen vertaisarviointiin.

Ensimmäisessä osatutkimuksessa tarkastelin oppilaiden toimintaa yhden vertaisarvioinnin aikana. Havaitsin, että rakentavan kriittisen palautteen saaminen oli oppilaille hyödyllistä, sillä se auttoi heitä kehittämään alkuperäistä työtään. Kuitenkin jopa vertaispalautteen vaikutusta merkityksellisempää oli oppilaan oma asennoituminen alkuperäiseen työhön ja hänen ymmärryksensä formatiivisen vertaisarvioinnin tarkoituksesta. Nämä tekijät liittyvät läheisesti oppilaan *palauteosaamiseen* (englanniksi *feedback literacy*, katso Carless & Boud, 2018), joka kuvaa asenteita ja taitoja, joita vaaditaan palautteen antajalta ja vastaanottajalta.

Tutkimukseni toisessa osassa tutkin ja vertasin oppilaiden palauteosaimista kahden samantapaisen vertaisarvioinnin yhteydessä ensin seitsemännen luokan puolella välissä ja vuotta myöhemmin kahdeksannella luokalla. Aineistosta erotin kolme erilaista palauteosaamisen lajia, joita kuvasin aineistosta tunnistetuilla osaamistasoilla. Tuloksista käy ilmi, että oppilaiden palauteosaaminen oli heterogeenista. Osa oppilaista ei ollut kiinnostunut palautteesta, kun taas jotkut etsivät ja pohtivat palautetta oma-aloitteisesti ja hyödynsivät sitä aktiivisesti. Lisäksi havaitsin, että oppilaiden palauteosaaminen kehittyi selvästi vuoden aikana, mahdollisesti juuri vertaisarvioinnin seurauksena. Palauteosaamista on aiemmin tutkittu lähinnä kolmannella asteella, mutta tulosten perusteella on selvää, että paljon nuoremmatkin oppilaat pystyvät kehittämään näitä taitoja.

Tutkimukseni kolmannessa osassa tutkin oppilaiden toimintaa sosiokulttuurisesta näkökulmasta tarkastelemalla heidän toimijuuttaan vertaisarvioinnin aikana eli tarkastelin heidän kykyään ja tapojaan toteuttaa vertaisarviointia luokahuoneen sosiaalisessa ympäristössä. Kuuden yksittäisen vertaisarvioinnin analyysi paljasti erilaisia toimijuuden muotoja, joita oppilaat toteuttivat kolmessa vertaisarviointien tarjoamassa asemassa: ryhmän jäsenenä, arvioijana ja arvioitavana. Toimijuuden muotojen lähempi tarkastelu osoitti, että näennäisesti

yhdenmukainen vaatimus osallistua vertaisarviointiin ei ollutkaan käytännössä yhdenmukainen. Tietyt rakentavan toimijuuden muodot, kuten toisten auttaminen ja kritisoiminen, olivat joidenkin oppilaiden tavanomaista toimintaa, kun taas toisilta ne vaativat oman totutun roolin ylittämistä. Kuitenkin havaitsin, että opettajan ja luokkakavereiden tuki auttoi oppilaita toteuttamaan myös sellaisia toimijuuden muotoja, jotka olivat heille haastavia.

Tutkimukseni tuloksia voidaan hyödyntää käytännön opetustyössä. Ensinäkin opettajat ovat valmiimpia toteuttamaan vertaisarviointia ja tukemaan oppilaita yksilöllisissä haasteissa, kun he ovat tietoisia oppilaiden palauteosaamisen kirjosta. Tutkimukseni tuottamia palauteosaamisen kuvauksia voidaan käyttää esimerkiksi oppilaiden vertaisarviointikoulutuksessa ja itsearviointin kriteerein. Lisäksi niiden avulla voidaan keskustella asennoitumisesta palautteeseen yleisemminkin. Jos opettajalle selviää, että oppilas ei halua palautetta, hänen voi olla tarkoituksenmukaista jättää palaute antamatta ja keskittyä sen sijaan oppilaan kanssa tavoitteen asetteluun: löytyisikö jokin tavoite, jonka saavuttamisesta oppilas haluaisi palautetta? Vertaisarviointi tarjoaa erinomaisen mahdollisuuden puhua arvioinnista metatasolla, esimerkiksi pohtia juuri palauteosaamisen taitoja. Tällaiset metakeskustelut ovat yksi syy, jonka vuoksi vertaisarviointi kehittää palauteosaamista.

Tutkimuksen toinen käytännön näkökulma liittyy oppilaiden toimijuuteen. Tieto siitä, että vertaisarviointi vaatii tietojen, taitojen ja asenteiden lisäksi toimijuutta, auttaa opettajia tukemaan oppilaita tarkoituksenmukaisesti. Väärin kohdistettu tuki saattaa jopa heikentää oppilaan toimijuutta. Jos opettaja tukee oppilaan suoritusta liian vahvasti ja esimerkiksi kertoo, miten työ tulisi arvioida tai mitä palautteeseen kannattaisi kirjoittaa, hän saattaa tulla viestineeksi, ettei usko oppilaan pystyvän tehtävään yksin. Opettajan kannattaa enemmän rohkaista oppilasta, tukea tämän ajatuksia (myös muiden kuullen) ja mahdollistaa ongelmista keskusteleminen toisten oppilaiden kanssa. Näin oppilas saa vahvistusta ajatuksilleen ja voi ikään kuin turvautua muiden toimijuuteen vertaisarvioinnista suoriutuakseen.

Tutkimukseni tulokset tuovat uutta tietoa vertaisarvioinnin vähemmän tunnetuista osa-alueista eli sen kytköksistä oppimaan oppimisen taitoihin sekä sosiokulttuurisiin tekijöihin. Vertaisarvioinnin avulla oppilaat voivat oppia arvostamaan ja hyödyntämään palautetta sekä harjoitella toimimista rakentavissa rooleissa. Palauteosaamisen ja toimijuuden vahvistaminen auttaa oppilaita olemaan aktiivisia oppimisessaan. Oppilaan arviointi on perinteisesti ollut summatiivisesti painottunutta, ja se on nähty opettajan vastuuna. Edelleen monet koulun käytännöt pitävät yllä tätä perinnettä, mikä osaltaan selittää oppilaiden passiivista suhtautumista arviointiin ja palautteeseen. Opettajan palautetta ei yleensä kannusteta kyseenalaistamaan, eikä summatiivista arviointia välttämättä käytetä oppimisen edistämiseen. Formatiivinen vertaisarviointi muuttaa asetelmaa, sillä sen aikana oppilaat ovat pääosassa ja opettajat tukevat oppilaiden välistä palautteen antamista ja vastaanottamista.

Tutkimukseni tulosten mukaan vertaisarviointi voi vahvistaa oppilaskeskistä oppimiskulttuuria, joka keskittyy ulkoisten tavoitteiden sijasta oppimiseen ja kannustaa yhteistyöhön. Tämä vaatii kuitenkin pitkäjänteistä työtä ja edellyttää paitsi koko oppilasryhmän ohjaamista myös oppilaiden tukemista yksilötasolla.

REFERENCES

- Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology*, 30, 109–137. <https://doi.org/10.1146/annurev.anthro.30.1.109>
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189x11428813>
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32(2), 159–181. <https://doi.org/10.1080/02602930600801928>
- 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. <https://doi.org/10.1080/02635143.2018.1557628>
- Ashton, S., & Davies, R. S. (2015). Using scaffolded rubrics to improve peer assessment in a MOOC writing course. *Distance Education*, 36(3), 312–334. <https://doi.org/10.1080/01587919.2015.1081733>
- Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., Tamm, M., Kamppi, P., Rumpu, N., Hietala, R., & Immonen, J. (2019). "Että tietää missä on menossa." *Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa*. The Finnish Education Evaluation Centre.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594x.2010.513678>
- Biesta, G., & Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults* 39(2), 132–149. <https://doi.org/10.1080/02660830.2007.11661545>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university*. Open University Press.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–44. <https://doi.org/10.1177/003172171009200119>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594x.2018.1441807>

- Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: A practical guide*. McGraw-Hill International.
- Boud, D. (2014). Shifting views of assessment: From secret teachers' business to sustaining learning. In C. Kreber, C. Anderson, N. Entwistle, & J. MacArthur (Eds.), *Advances and innovations in university assessment and feedback* (pp. 13–31). Edinburgh University Press.
<https://doi.org/10.3366/edinburgh/9780748694549.003.0002>
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 31(4), 399–413.
<https://doi.org/10.1080/02602930600679050>
- Boud, D., & Falchikov, N. (2007). Introduction: assessment for the longer term. In D. Boud, & N. Falchikov (Eds.), *Rethinking Assessment for Higher Education: Learning for the Longer Term*. (pp. 3–13). Routledge.
<https://doi.org/10.4324/9780203964309>
- Boud, D., & Molloy, E. (2013). What is the problem with feedback? In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding it and doing it well* (pp. 1–10). Routledge.
- Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
<https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., Clarke, V., & Weate, P. (2016). Using thematic analysis in sport and exercise research. In B. Smith & A. C. Sparkes (Eds.), *Routledge handbook of qualitative research in sport and exercise* (pp. 191–205). Routledge.
- Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22(52), 782–861.
<https://doi.org/10.1111/j.1468-0327.2007.00189.x>
- Cambridge Dictionary. (n.d.). Citation. In Cambridge Dictionary. Retrieved February 25, 2021 from <https://dictionary.cambridge.org/>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education*, 42(6), 930–941.
<https://doi.org/10.1080/02602938.2016.1211246>
- Carless, D., & Winstone, N. (2020): Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*. Advance online publication. <https://doi.org/10.1080/13562517.2020.1782372>
- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, 35(5), 551–564.
<https://doi.org/10.1080/02602931003632381>
- Chang, H.-Y., & Quintana, C., & Krajcik, J. (2009). The Impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education*, 94(1), 73–94.
<https://doi.org/10.1002/sce.20352>

- Charteris, J., & Thomas, E. (2017). Uncovering 'unwelcome truths' through student voice: Teacher inquiry into agency and student assessment literacy. *Teaching Education*, 28(2), 162–177.
<https://doi.org/10.1080/10476210.2016.1229291>
- Chetcuti, D., & Cutajar, C. (2014). Implementing peer assessment in a post-secondary (16–18) physics classroom. *International Journal of Science Education*, 36(18), 3101–3124.
<https://doi.org/10.1080/09500693.2014.953621>
- Chong, S. W. (2019). The use of exemplars in English writing classrooms: from theory to practice. *Assessment & Evaluation in Higher Education*, 44(5), 748–763. <https://doi.org/10.1080/02602938.2018.1535051>
- Chong, S. W. (2020). Reconsidering student feedback literacy from an ecological perspective. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2020.1730765>
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205–249.
<https://doi.org/10.1007/s10648-011-9191-6>
- Connolly, L. M. (2016). Trustworthiness in Qualitative Research. *Medsurg Nursing*, 25(6), 435–436.
- Cope, D.G. (2014). Methods and meanings: Credibility and trustworthiness of qualitative research. *Oncology Nursing Forum*, 41(1), 89–91.
<https://doi.org/10.1188/14.onf.89-91>
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice*, 6(1), 101–116. <https://doi.org/10.1080/09695949993026>
- Crane, L., & Winterbottom, M. (2008). Plants and photosynthesis: Peer assessment to help students learn. *Journal of Biological Education*, 42(4), 150–156. <https://doi.org/10.1080/00219266.2008.9656133>
- Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *The Phi Delta Kappan*, 73(3), 220–225.
- Darling-Hammond, L., Austin, K., Orcutt, S., Rosso, J. (2001). How people learn: Introduction to learning theories. In *The learning classroom: Theory into practice a telecourse for teacher education and professional development* (pp. 1–22). Stanford University.
<http://web.stanford.edu/class/ed269/hplintrochapter.pdf>. Accessed 24 September, 2020.
- Davis, N., Kumtepe, E. G., & Aydeniz, M. (2007). Fostering continuous improvement and learning through peer assessment: Part of an integral model of assessment. *Educational Assessment*, 12(2), 113–135.
<https://doi.org/10.1080/10627190701232720>
- Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, 44(1), 25–36.
<https://doi.org/10.1080/02602938.2018.1467877>

- Delva, D., Sargeant, J., Miller, S., Holland, J., Brown, P. A., Leblanc, C., Lightfoot, K., & Mann, K. (2013). Encouraging residents to seek feedback. *Medical Teacher*, 35(12), 1625–1631. <https://doi.org/10.3109/0142159x.2013.806791>
- Dolezal, D., Posekany, A., Roschger, C., Koppensteiner, G., Motschnig, R., & Pucher, R. (2018). Person-centered learning using peer review method: An evaluation and a concept for student-centered classrooms. *International Journal of Engineering Pedagogy*, 8(1), 127–147. <https://doi.org/10.3991/ijep.v8i1.8099>
- Double, K. S., McGrane, J. A., & Hopfenbeck, T.N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32, 481–509. <https://doi.org/10.1007/s10648-019-09510-3>
- Edwards, A. (2009). From the systemic to the relational: Relational agency and activity theory. In A. Sannino, H. Daniels, & K. D. Gutierrez (Eds.), *Learning and expanding with activity theory* (pp. 197–211). Cambridge University Press. <https://doi.org/10.1017/cbo9780511809989.013>
- Edwards, A., & D'Arcy, C. (2004). Relational agency and disposition in sociocultural accounts of learning to teach. *Educational Review*, 56(2), 147–155. <https://doi.org/10.1080/0031910410001693236>
- Egelandsdal, K., & Riese, H. (2020). Never mind the gap: Formative assessment confronted with Dewey's and Gadamer's concept of experience. *European Journal of Education*, 55(1), 91–104. <https://doi.org/10.1111/ejed.12378>
- Elwood, J. (2006). Gender Issues in testing and Assessment. In C. Skelton, B. Francis, and L. Sumyan (Eds.), *The SAGE Handbook of Gender and Education* (pp. 262–272). SAGE. <https://doi.org/10.4135/9781848607996.n20>
- Elwood, J., & Murphy, P. (2015). Assessment systems as cultural scripts: a sociocultural theoretical lens on assessment practice and products. *Assessment in Education Principles Policy and Practice*, 22(2), 182–192. <https://doi.org/10.1080/0969594x.2015.1021568>
- Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, 103(4), 962–1023. <https://doi.org/10.1086/231294>
- Eteläpelto, A., Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2013). What is agency? Conceptualizing professional agency at work. *Educational Research Review*, 10, 45–64. <https://doi.org/10.1016/j.edurev.2013.05.001>
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287–322. <https://doi.org/10.3102/00346543070003287>
- Finnish National Board of Education. (1999). *Perusopetuksen oppilaan arvioinnin perusteet 1999* [The Fundamentals of Student Evaluation in Basic Education 1999]. Opetushallitus.
- Finnish National Board of Education. (2014). *Perusopetuksen opetussuunnitelman perusteet 2014* [National Core Curriculum for Basic Education 2014]. Opetushallitus.

- Finnish National Board of Education. (2020). *Oppilaan oppimisen ja osaamisen arviointi perusopetuksessa*. [Assessment of pupils' learning and competence in basic education]. Opetushallitus.
- Finnish Advisory Board of Research Integrity. (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. Finnish Advisory Board of Research Integrity. https://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf. Accessed 21 September, 2020.
- Foley, S. (2013). Student views of peer assessment at the International School of Lausanne. *Journal of Research in International Education*, 12(3), 201–213. <https://doi.org/10.1177/1475240913509766>
- Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 42(6), 861–878. <https://doi.org/10.1007/s11251-014-9319-4>
- Gielen, 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. <https://doi.org/10.1016/j.learninstruc.2009.08.007>
- Gipps, C. (1999). Chapter 10: Socio-Cultural Aspects of Assessment. *Review of Research in Education*, 24(1), 355–392. <https://doi.org/10.3102/0091732x024001355>
- Glaserfeld, E. v. (1992). A constructivist's view of learning and teaching. In Duit, R., Goldberg, F., & Niedderer, H. (Eds). *Research in Physics Learning: Theo-retical Issues and Empirical Studies : Proceedings of an international workshop held at the University of Bremen, March 4-8, 1991*. (pp. 29–39). Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel
- Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, 8(4), 597–606. <https://doi.org/10.46743/2160-3715/2003.1870>
- Gresalfi, M., Martin, T., Hand, V., & Greeno, J. (2009). Constructing competence: An analysis of student participation in the activity systems of mathematics classrooms. *Educational Studies in Mathematics*, 70(1), 49–70. <https://doi.org/10.1007/s10649-008-9141-5>
- Hafen, C. A., Hamre, B., Allen, J. P., Courtney, A. B., Gitomer, D. H., & Pianta, R. C. (2015). Teaching through interactions in secondary school classrooms: Revisiting the factor structure and practical application of the classroom assessment scoring system – secondary. *Journal of Early Adolescence*, 35(5–6), 651–680. <https://doi.org/10.1177/0272431614537117>
- Haimovitz, K., & Dweck, C. S. (2017). The origins of children's growth and fixed mindsets: New research and a new proposal. *Child Development*, 88(6), 1849–1859. <https://doi.org/10.1111/cdev.12955>
- Han, Y., & Xu, Y. (2019a). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education*. Advance online publication. <https://doi.org/10.1080/13562517.2019.1648410>

- Han, Y., & Xu, Y. (2019b). The development of student feedback literacy: The influences of teacher feedback on peer feedback. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2019.1689545>
- Harris, L., & Brown, G. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies in teachers' implementation. *Teaching and Teacher Education, 36*, 101–111. <https://doi.org/10.1016/j.tate.2013.07.008>
- Harris, L. R., Brown G. T., & Dargusch J. (2018). Not playing the game: Student assessment resistance as a form of agency. *The Australian Educational Researcher, 45*(1), 125–140. <https://doi.org/10.1007/s13384-018-0264-0>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hey-Cunningham, A. J., Ward, M-H., & Miller, E. J. (2020). Making the most of feedback for academic writing development in postgraduate research: Pilot of a combined programme for students and supervisors. *Innovations in Education and Teaching*. Advance online publication. <https://doi.org/10.1080/14703297.2020.1714472>
- Hilden, R., Ouakrim-Soivio, N., & Rautopuro, J. (2016). Kaikille ansionsa mukaan? Perusopetuksen päättöarviointin yhdenvertaisuus Suomessa. *Kasvatus, 47*(4), 342–357.
- 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. <https://doi.org/10.1016/j.compedu.2013.09.019>
- Hsia, L. - H., Huang, I., & Hwang, G.-J. (2015). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses: Peer assessment for performing art courses. *British Journal of Educational Technology, 47*(4), 618–632. <https://doi.org/10.1111/bjet.12248>
- Johnson, N., & Winterbottom, M. (2011). Supporting girls' motivation in science: a study of peer- and self-assessment in a girls-only class. *Educational Studies, 37*(4), 391–403. <https://doi.org/10.1080/03055698.2010.508605>
- Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education, 14*(1), 63–76. <https://doi.org/10.1177/1469787412467125>
- Kim, H., & Song, J. (2006). The features of peer argumentation in middle school students' scientific inquiry. *Research in Science Education, 36*(3), 211–233. <https://doi.org/10.1007/s11165-005-9005-2>
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>

- Kumpulainen, K., & Rajala, A. (2017). Dialogic teaching and students' discursive identity negotiation in the learning of science. *Learning and Instruction, 48*, 23–31. <https://doi.org/10.1016/j.learninstruc.2016.05.002>
- Kurihara, N. (2017). Do peer reviews help improve student writing abilities in an EFL high school classroom? *TESOL Journal, 8*(2), 450–470. <https://doi.org/10.1002/tesj.282>
- Le Hebel, F. (2017). *Peer-assessment effectiveness in developing students' understanding*. ESERA 2017.
- Lepak, J. (2014). Enhancing students' written mathematical arguments. *Mathematics Teaching in the Middle School, 20*(4), 212–219. <https://doi.org/10.5951/mathteacmidscho.20.4.0212>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of PA. *Teaching in Higher Education, 11*(3), 279–290. <https://doi.org/10.1080/13562510600680582>
- 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. <https://doi.org/10.1007/s11251-011-9177-2>
- Luukkonen, V. (2020). *Student's perspectives on the student-teacher-parent discussion in comprehensive school in Finland*. [Doctoral dissertation, University of Jyväskylä]. <https://jyx.jyu.fi/handle/123456789/73155>
- Matusov, E. (2011). Authorial teaching and learning. In E. J. White & M. Peters (Eds.), *Bakhtinian pedagogy: Opportunities and challenges for research, policy and practice in education across the globe* (pp. 21–46). Peter Lang Publishers.
- Matusov, E., von Duyke, K., & Kayumova, S. (2016). Mapping concepts of agency in educational contexts. *Integrative Psychological and Behavioral Science, 50*(3), 420–446. <https://doi.org/10.1007/s12124-015-9336-0>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed). SAGE.
- Minjeong, K. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education, 34*(1), 105–114. <https://doi.org/10.1080/02602930801955960>
- Mok, J. (2011). A case study of students' perceptions of peer assessment in Hong Kong. *ELT Journal, 65*(3), 230–239. <https://doi.org/10.1093/elt/ccq062>
- Molloy, E., Boud, D., & Henderson, M. (2019). Developing a learning-centered framework for feedback literacy. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2019.1667955>
- Nash, R. A., & Winstone, N. E. (2017). Responsibility-sharing in the giving and receiving of assessment feedback. *Frontiers in Psychology, 8*. <https://doi.org/10.3389/fpsyg.2017.01519>

- National Advisory Board on Research Ethics. (2009). *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review*. <http://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf>.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education* 39(1), 102–122.
<https://doi.org/10.1080/02602938.2013.795518>
- Nieminen, J. H. (2020). *The lunatics have taken over the assessment: Utilising self-assessment to theorise - and disrupt - the interplay of agency and power in undergraduate mathematics education*. [Doctoral dissertation, University of Helsinki]. <http://urn.fi/URN:ISBN:978-951-51-6366-0>
- Nieminen, J. H., & Tuohilampi, L. (2020). 'Finally studying for myself' – Examining student agency in summative and formative self-assessment models. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2020.1720595>
- Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence-Based Nursing*, 18(2), 34–35.
<https://doi.org/10.1136/eb-2015-102054>
- Norris, N. (1997). Error, bias and validity in qualitative research. *Educational Action Research*, 5(1), 172–176.
<https://doi.org/10.1080/09650799700200020>
- Nuthall, G. (2005). The cultural myths and realities of classroom teaching and learning: A personal journey. *Teachers College Record*, 107(5), 895–934.
<https://doi.org/10.1111/j.1467-9620.2005.00498.x>
- O'Donovan, B. (2017). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education*, 74(4), 617–633. <https://doi.org/10.1007/s10734-016-0068-y>
- Oinas, S. (2020). *Technology-enhanced feedback: Teachers' practices, pupils' perceptions and their relations to learning and academic well-being*. [Doctoral dissertation, University of Helsinki]. <http://urn.fi/URN:ISBN:978-951-51-6733-0>
- Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity: International Journal of Methodology*, 41(2), 233–249. <https://doi.org/10.1007/s11135-006-9000-3>
- Ouakrim-Soivio, N. (2013). *Toimivatko päättöarvioinnin kriteerit? Oppilaiden saamat arvot ja opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina*. [Doctoral dissertation, University of Helsinki]. <http://urn.fi/URN:ISBN:978-952-13-5579-0>
- Ouakrim-Soivio, N., Rautopuro, J., & Saario, J. (2017). Yhdenvertaisuutta vai armopaloja? Sivistyksellisen yhdenvertaisuuden toteutuminen perusopetuksessa. *Kasvatus*, 48(3), 174–188.
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp. 247–266). Routledge.

- Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133–156. <https://doi.org/10.1007/s10212-015-0282-5>
- Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know? In A.A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 409–431). Cambridge University Press.
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on construct validity of peer assessment, perceived fairness and comfort, and performance. *Studies In Educational Evaluation*, 39(4), 195–203. <https://doi.org/10.1016/j.stueduc.2013.10.005>
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238–250. <https://doi.org/10.1016/j.learninstruc.2007.05.001>
- Rainio, A. P. (2008). From resistance to involvement: Examining agency and control in a playworld activity. *Mind, Culture, and Activity*, 15(2), 115–140.
- Rajala, A. (2016). *Toward an agency-centered pedagogy A teacher's journey of expanding the context of school learning*. [Doctoral Dissertation, University of Helsinki] <http://urn.fi/URN:ISBN:978-951-51-2252-0>
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4–13. <https://doi.org/10.1002/bs.3830280103>
- Reeves, T. C. (2006). How do you know they are learning? The importance of alignment in higher education. *International Journal of Learning Technology*, 2(4), 294–309. <https://doi.org/10.1504/ijlt.2006.011336>
- Reimann, N., Sadler, I., & Sambell, K. (2019). What's in a word? Practices associated with 'feedforward' in higher education. *Assessment & Evaluation in Higher Education*, 44(8), 1279–1290. <https://doi.org/10.1080/02602938.2019.1600655>
- Rowley, J. (2002). Using case studies in research. *Management Research News* 25(1), 16–27. <https://doi.org/10.1108/01409170210782990>
- 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). Multilingual Matters. <https://doi.org/10.21832/9781847697912-010>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/bf00117714>
- Silseth, K., & Gilje, Ø. (2019). Multimodal composition and assessment: a sociocultural perspective. *Assessment in Education: Principles, Policy & Practice*, 26(1), 26–42. <https://doi.org/10.1080/0969594x.2017.1297292>
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment, the training of peer-assessment skills*. Interuniversity Centre for Educational Research.
- Smith, C. D., Worsfold, K., Davies, L. Fisher, R., & McPhail, R. (2013) Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy', *Assessment & Evaluation in*

- Higher Education*, 38(1), 44–60.
<http://dx.doi.org/10.1080/02602938.2011.598636>
- Stutchbury, K., & Fox, A. (2009). Ethics in educational research: introducing a methodological tool for effective ethical analysis. *Cambridge Journal of Education*, 39(4), 489–504. <https://doi.org/10.1080/03057640903354396>
- Sutton, P. (2012). Conceptualizing feedback literacy: knowing, being, and acting. *Innovations in Education and Teaching International* 49(1), 31–40. <https://doi.org/10.1080/14703297.2012.647781>
- Tasker, T., & Herrenkohl, L. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education*, 27(1), 35–59. <https://doi.org/10.1007/s10972-016-9454-7>
- Thomas, G. (2021). *How To Do Your Case Study* (3rd ed.). SAGE.
- To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2018.1548559>
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. <https://doi.org/10.3102/00346543068003249>
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20–27. <https://doi.org/10.1080/00405840802577569>
- 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). SAGE. <https://doi.org/10.4135/9781452218649.n22>
- Topping, K. J. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, 1(1), 1–17. <https://doi.org/10.31532/interdiscipeducpsychol.1.1.007>
- Tsagari, D., & Meletiadou, E. (2015). Peer assessment of adolescent learners' writing performance. *Writing & Pedagogy*, 7(2), 305–328. <https://doi.org/10.1558/wap.v7i2-3.26457>
- 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(4), 1161–1174. <https://doi.org/10.1016/j.compedu.2006.01.007>
- 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. <https://doi.org/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(4), 506–519. <https://doi.org/10.1016/j.learninstruc.2010.08.002>
- 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.
- Wang, X., Hwang, G., Liang, Z., & Wang, H. (2017). Enhancing students' computer programming performances, critical thinking awareness and attitudes towards programming: An online peer-assessment attempt. *Journal of Educational Technology & Society*, 20(4), 58–68.
<https://www.jstor.org/stable/26229205>
- Wiliam, D. (2012). Feedback: Part of a system. *Educational Leadership*, 70(1), 30–34.
- Winstone, N. E. & Boud, D. (2020). The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*. Advance online publication. <https://doi.org/10.1080/03075079.2020.1779687>
- Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). It'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience. *Studies in Higher Education* 42(11), 2026–2041.
<https://doi.org/10.1080/03075079.2015.1130032>
- Winstone, N., Pitt, E., & Nash, R. (2020). Educators' perceptions of responsibility-sharing in feedback processes. *Assessment & Evaluation in Higher Education*. Advance online publication.
<https://doi.org/10.1080/02602938.2020.1748569>
- Wisniewski, B., Zierer, K., Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10.
<https://doi.org/10.3389/fpsyg.2019.03087>
- Vygotsky, L. S. (1978). In M. Cole, V. John-Steiner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Won, M., Krabbe, H., Ley, S. L., Treagust, D. F., & Fischer, H. E. (2017). Science teachers' use of a concept map marking guide as a formative assessment tool for the concept of energy. *Educational Assessment*, 22(2), 95–110.
<https://doi.org/10.1080/10627197.2017.1309277>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162.
<https://doi.org/10.1016/j.tate.2016.05.010>
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). SAGE.
- 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.
<https://doi.org/10.1080/13803611.2014.964261>
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41–54.
<https://doi.org/10.1016/j.edurev.2008.11.002>
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and

conceptions. *Learning and Instruction*, 20(4), 280–290.

<https://doi.org/10.1016/j.learninstruc.2009.08.010>

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

<https://doi.org/10.1016/j.learninstruc.2009.08.004>

APPENDICES

Appendix 1. The Description and Material of Peer Assessment Intervention

Training Session 1: Reflecting on the purpose of assessment

The goal of this session was to make the students aware of the summative and formative purposes of assessment.

Implementation

The students were given question sheets (Figure 1) and time to individually reflect and write down as many purposes for assessment as they could think of.

Answers were discussed in pairs or small groups.

The researcher asked the students to explain their views in a class discussion.

Answers were written on the blackboard, which had been divided in half: “Helps learning” and “Measures learning.” The researcher wrote the students’ answers on the appropriate side (or in the middle). As there were no right or wrong answers, all views were welcome.

In the end, the researcher explained that formative assessment is important for learning.

Name: _____
<p>What are the uses of assessment at school? Write down as many as you can think of.</p>

Figure 1: A model of the question sheet.

Training Session 2: Reflecting on the usefulness of assessment

The goal of this session was to teach the students how to consider useful elements of feedback and understand that feedback is meant to help them learn.

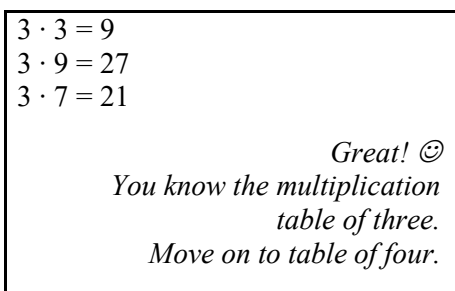
Name: _____
<p>What kind of assessment helps you? What kind of assessment advances your learning? Write down as many things as you can think of.</p>

Figure 2: A model of the question sheet.

Implementation

1. The students were given question sheets (Figure 2) and time to individually reflect and write down as many useful forms of assessment as they could think of.
2. The researcher explained that the purpose of feedback is to help students achieve their goals. If it does not do that, it is not useful.
3. The researcher presented examples of feedback (Figure 3) and asked the students for their views on the feedback's usefulness. All views were welcome, even though the cards were intended to point out certain themes.

Card	Points of the cards
$1 + 3 = 5$ $4 + 3 = 8$ <p style="text-align: right;"><i>Well done!! Keep up the good work!</i></p>	<p>The most important criterion for feedback is that it is <i>accurate</i>.</p>
$1 + 3 = 13 \quad \textit{Wrong!}$ $4 + 3 = 43 \quad \textit{Wrong!}$ $2 + 9 = 29 \quad \textit{Wrong!}$	<p>Another criterion for feedback is that it helps the student <i>move forward</i>.</p> <ul style="list-style-type: none"> • “Wrong” does not help the student move forward. • Such feedback may even lower motivation. • The student’s work shows a pattern of thinking. Why is it wrong? How should the equations be calculated?
$\frac{3^4 \cdot 9^2}{3(127 - 4 \cdot 31)} = \frac{3^8}{3 \cdot 3}$ $= 3^6 = 729$ <p style="text-align: right;"><i>This was innovative! I do not often see work this skillful.</i></p>	<p>Feedback may help someone by increasing motivation or self-efficacy. This feedback might help with that, but it has no other use.</p>
$1 + 1 = 2$ <p style="text-align: right;"><i>Great!!! This was a top performance!</i></p>	<p>This is similar to the example above.</p> <p>Does the feedback raise or lower self-confidence? That depends on the student’s skills and goals.</p>



This feedback has good qualities:

- It presents what the student already knows.
- It presents what could be learned next.

Figure 3: Cards for classroom discussion.

Training Session 3: Creating criteria of scientific inquiry

This session's context was a project exploring the growth of peas. This was the first scientific inquiry that the students planned, conducted, and reported on. The goal of the training was to make them consider the criteria for the successful accomplishment of scientific inquiry?

Implementation

1. The teacher presented the task to the students and explained that they were not expected to write a report but to think about how the report should be assessed. The instructions for the task were:
 - How should an inquiry into the growth of peas be assessed? What should one receive feedback on?
 - Plan six criteria to use for assessment. Keep in mind that the work is well done if a student achieves all six.
2. The teacher offered an example of someone brushing their teeth and explained the assessment criteria (Figure 4). The purpose of the example criteria was to help the students get started but not to give direct answers. The word "criteria" was potentially unfamiliar, and the example clarified it.
3. The teacher delivered blank assessment rubrics (Figure 5). The students were allowed to work in pairs or groups, but everyone had to fill out their own paper.
4. A class discussion followed. The teacher and the researcher asked for proposed criteria, and these were discussed. The students were still allowed to change their own criteria if they wished.
5. The teacher collected the sheets for later use.

Criteria	Points
The toothbrush was soft enough.	0 ½ 1
The toothpaste contained fluoride.	0 ½ 1
The toothbrush was less than three months old.	0 ½ 1
The person was brushing inside surfaces, outside surfaces, and chewing surfaces.	0 ½ 1
The person brushed their teeth for two minutes.	0 ½ 1
The person spit out toothpaste when they were finished.	0 ½ 1

Figure 4: Example criteria for assessing someone's brushing of their teeth.

Criteria	Points		
	0	½	1
	0	½	1
	0	½	1
	0	½	1
	0	½	1
	0	½	1

Figure 5: The students' blank assessment rubric.

Training Session 4: Assessing a report using self-made criteria

This session's context was the same pea project used in Training Session 3. The goal was to practice assessing and to continue processing the criteria. For this task, the researcher had prepared a fictional report on growing peas.

Implementation

1. The teacher explained that the goal of the task was to train the students to be assessors and explained the purpose of the three-step assessment rubric using the example of a person brushing their teeth (Figure 4).
2. The students were given the criteria that they had created in Training Session 3 as well as a copy of a report by a fictional student.
3. Everyone assessed the report according to their own criteria.
4. Based on the points immediately below, the teacher and the researcher led a class discussion about assessing.
 - How did assessing feel?
 - Was it easy to decide which number to choose?
 - How did the criteria work?
 - What worked well?
 - What did not?
5. A moment was taken to consider the reliability of the assessment. The researcher asked the students to add up the points and write the sum on the blackboard. The results were compared, and students noticed that they varied. The students concluded that assessment is difficult and that the choice of criteria is crucial.

Training Session 5: Self-assessment of scientific inquiry

This session's context was the pea growing project. The goal was to practice assessing work against criteria.

Implementation

1. The teacher presented eight criteria for students' work (Figure 6). She explained that the students were supposed to compare their own work to the criteria and mark it accordingly: 0 = not accomplished, $\frac{1}{2}$ = partly accomplished, or 1 = fully accomplished.
2. The students were given the criteria, and everyone assessed their own work.
3. The assessments were discussed in pairs or small groups. The teacher and the researcher listened to the group discussions and participated when appropriate.

Criteria	Points		
The research plan explains how the conditions for this crop of peas differed from the comparison crop.	0	$\frac{1}{2}$	1
The report presents a hypothesis and justifies it.	0	$\frac{1}{2}$	1
The equipment is clearly listed.	0	$\frac{1}{2}$	1
The peas have been grown according to the plan.	0	$\frac{1}{2}$	1
Measurements are clearly reported.	0	$\frac{1}{2}$	1
The suitability of the conditions for growing peas has been considered.	0	$\frac{1}{2}$	1
Error of measurements have been noted.	0	$\frac{1}{2}$	1
There are suggestions as to how the inquiry can be improved.	0	$\frac{1}{2}$	1

Figure 6: Self-assessment rubric provided by the teacher.

Training Session 6: Reflecting on the quality of feedback and rules of peer assessment

This training session was a warm-up for the first actual peer assessment. The goal was to remind the students of the qualities of good feedback and present the rules of peer assessment.

Implementation

1. The researcher introduced examples of vague versus detailed feedback using the context of a person brushing their teeth. The class then discussed the helpfulness of the feedback (Figure 7).

You brushed quite poorly.	You only brushed chewing surfaces, but there are bacteria on inside and outside surfaces as well.
---------------------------	---------------------------------------------------------------------------------------------------

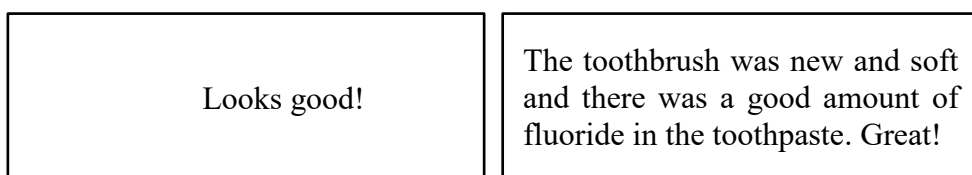


Figure 7: Example cards demonstrating vague and detailed feedback.

2. The researcher reminded the students that giving and receiving feedback is not easy—they need practice. Cards showing the rules of peer assessment (Figure 8) were presented and discussed.

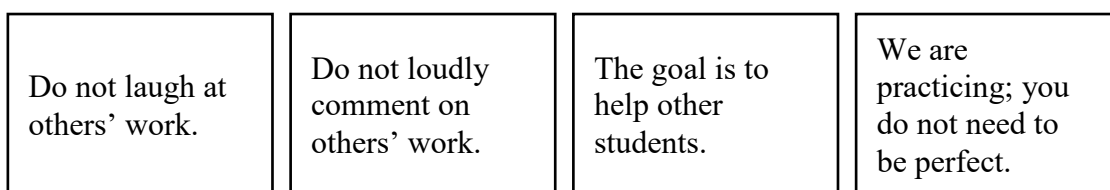


Figure 8: Cards showing peer assessment rules.

Peer Assessment Session 1: Groups assessing project plans

This session's (PA1) context was a technology project for which groups planned and built a model of a rover that could move on its own. The goal of the session was for the students to gain experience giving and receiving feedback and for them reconsider their project plan.

Implementation

1. After the plans were ready, the students were provided with the assessment criteria immediately below (they were discussed with the teacher and left on the whiteboard during the session).
 - The structure of the rover is strong enough to endure use several times.
 - It is possible to build the rover in a period of 1.5 hours.
 - The rover moves for long and far enough to make measuring time and distance possible.
2. The groups' project plans were spread out in different areas of the classroom, and groups visited every area to provide feedback. There was one group in each area, and they moved to the next area when the teacher gave a signal.
3. The students wrote comments, questions, and hints on Post-its and attached them to each plan. Each group had a different color Post-it so the teacher and the researcher could track the comments if necessary.
4. The students returned to their own plans and read and discussed the feedback they received. After, they had time to improve their work.

Training Session 7: Strategies to react to feedback and Peer Assessment Session 2: Individuals assessing a physics report

This session's (PA2) context was a physics inquiry determining the speed of the rover's built in the previous session. The goal was for the students to practice peer assessment and, by assessing and receiving feedback, learn to make and report an inquiry. Feedback was provided and work revised in two lessons.

Implementation

1. The students finished their own work and returned it to the teacher.
2. As a motivation for peer assessment, the researcher used cards to remind the students why peer assessment is useful (Figure 9).

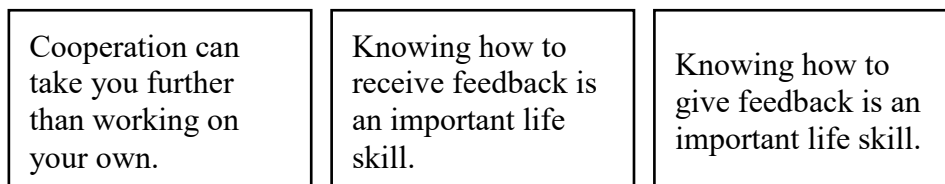


Figure 9: Rationale for peer assessment.

3. The teacher presented the instructions and assessment criteria and answered the students' questions.
4. The students were given a ready-made assessment rubric with criteria (Figure 11) and other student's work. The teacher and the researcher had already planned the pairs the students were divided into.
5. The students had 45 minutes to assess their peers' work. The teacher and the researcher circulated throughout the classroom and helped the students as needed.
6. The teacher collected the papers.
7. Before delivering feedback to the students, there was a warm-up activity about receiving feedback. The researcher presented the cards (Figure 10) one by one and asked the students for their opinions on the feedback.
8. After receiving peer feedback, the students had 45 minutes to improve their work. They then returned it to the teacher for summative assessment.

Card	Idea behind the card
Explain more about the concept of speed.	Do you think you have sufficiently explained the concept of speed? If not, add some information.
Bad handwriting.	Though this is not skillful feedback, can you learn something from it? Is good handwriting a criterion for this work? If not, do not waste too much energy on this. If your handwriting is bad, it might be a good idea to improve it in the future.
Lousy work. I could not understand anything.	This feedback is lousy. One cannot imagine it would be helpful. Do not waste time or energy on it but forget it as soon as possible. It is more reflective of the assessor than your work.

Figure 10: Example feedback cards.



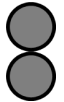
	Lacking 	Partly done 	Everything is ok 	Comment: What was good? What could be improved and how?
The equipment is clearly listed.				
The research plan is clearly presented.				
Measurements show the duration and distance the rover traveled.				
The results show how the speed was calculated from the measurements.				
The calculation includes an equation, numbers, and units.				
Error of measurements and results were explained.				




Figure 11: PA2's assessment rubric.

Peer Assessment Sessions 3: Members of pairs assess each other's lab work

The contexts of these peer assessment sessions (3.1, 3.2, 3.3, 3.4) were simple chemistry inquiries. For example, one task was to examine which substances dissolve in water. The goal of the sessions was to remind the students of the criteria for good lab work and to provide them with feedback on their work.

Implementation

1. After the teacher gave instructions for the inquiry, the researcher gave the students feedback forms that included the criteria for the task (Figure 12).
2. After finishing their lab work, the students assessed each other.
3. The students exchanged their assessments with their partner, read the feedback, and noted their agreement or disagreement. In most cases, they discussed the feedback with their partner, which was encouraged.
4. When the students were finished, they returned their papers to the researcher

Assessor	Needs improvement	Good work	Success!
Student being assessed			
You read the instructions before starting.			
You used the correct amounts of ingredients			
You understood what a solution looks like.			
You made precise observations.			
You wrote down the results.			
You cleaned up your work.			

Assessor writes

In space below, write down your lab partner's successes.

Student being assessed

Did you agree with the feedback?
(Circle the most suitable option.)

Strongly disagreed Disagreed Agreed Strongly agreed

Figure 12: An example of the assessment rubric from PA3.3

Peer Assessment Session 4: Individuals assessing a chemistry plan

This session's (PA4) context was a chemistry inquiry. The students had to plan how to separate a mixture of iron powder, salt, sand, and sawdust. The goal of the session was for them to practice assessing peers and, before they conducted the inquiry in groups, give them time to ponder the task on their own.

Implementation

The implementation was similar to PA2.

1. The students individually planned how the substances would be separated.
2. The students were given a ready-made assessment rubric (Figure 13) and other students' work. The teacher and the researcher had already planned the pairs the students were divided into.
3. The students had 10 minutes to assess peers' work. The teacher and the researcher circulated throughout the classroom and helped the students as needed.
4. The teacher collected the papers and returned them to the students.
5. After reading peer feedback, the students had 10 minutes to improve their work.

	Yes	No	Comment Were there problems in the plan? What could be improved and how? What worked well?
The substances were separated in the appropriate order.			
The measures in this plan will separate sand.			
The measures in this plan will separate sawdust.			
The measures in this plan will separate salt.			
The measures in this plan will separate iron dust.			
There is something wrong with the plan.			
I can learn something from this plan.			

Figure 13: PA4's assessment rubric

Training Session 8: Categorizing feedback

This session's goal was to remind students of the qualities of productive feedback and of the criteria for a physics inquiry. The session was adapted from Tasker and Herrenkohl (2016), but since the topics had been discussed earlier, it was arranged as a lighter version.

Implementation

The researcher had chosen 22 feedback comments (Figure 14) that the students had provided each other during PA2. The students worked in the groups of 3–4 and to followed the instructions below.

1. Divide the comments into different categories according to their quality.
2. Create names for the categories.
3. Write the category names on the cardboard, and classify the feedback according to the names.
4. If you face problems, for example a comment fits in several categories, use your creativity.

While the students worked, the teacher and the researcher circulated throughout the classroom and discussed various issues with the groups. At the end of the session, they discussed the task with each group.





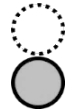
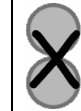
The research plan is presented clearly.				<i>You could have explained how the measurements were made a little more clearly.</i>
The research plan is presented clearly.				<i>Well done!</i>

Figure 14: Two examples of feedback comments.

Peer Assessment Session 5: Individuals assessing a physics report

This session's context (PA5) was a physics inquiry exploring the resistance of a wire and defining how its length affects the current in a circuit. The goal was to practice peer assessment and, by assessing and receiving feedback, learn about making and reporting on an inquiry.

Implementation

1. The students finished their own work and returned it to the teacher.
2. The teacher presented the instructions and assessment criteria and answered the students' questions.
3. The students were given a ready-made assessment rubric with criteria (Figure 15) and other students' work. The teacher and the researcher had already planned the pairs the students were divided into.
4. The students had 30 minutes to assess their peers' work. The teacher and the researcher circulated throughout the classroom and helped the students when needed.
5. The teacher collected the papers.
6. The students were given back their work and peer feedback, and they had 30 minutes to improve their work. After, they returned their work to the teacher for summative assessment.



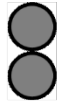
	Lacking 	Partly done 	Everything is ok 	Comment What was good? What could be improved and how?
The equipment is listed clearly. Another group could repeat the inquiry with the information.				
The research plan is explained clearly. It can be understood without already knowing the instructions.				
The results show the value of the current of several wire lengths.				
The results include a graph showing how the length of the wire affects the current.				
The axes are named, equally spaced, and have units.				
Error of measurements and results are carefully explained.				

Figure 15: PA5's assessment rubric.

References:

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



ORIGINAL PAPERS

I

PATHWAYS THROUGH PEER ASSESSMENT: IMPLEMENTING PEER ASSESSMENT IN A LOWER SECONDARY PHYSICS CLASSROOM

by

Laura Ketonen, Markus Hähkiöniemi, Pasi Nieminen & Jouni Viiri, 2020

International Journal of Science and Mathematics Education,
18, 1465-1484

<https://doi.org/10.1007/s10763-019-10030-3>

Reproduced with kind permission by Springer.



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

Laura Ketonen¹  · Markus Hähkiöniemi¹ · Pasi Nieminen¹ · Jouni Viiri¹

Received: 4 June 2019 / Accepted: 16 October 2019 / Published online: 2 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
Improving work	Not constructive	No constructive comments
	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.

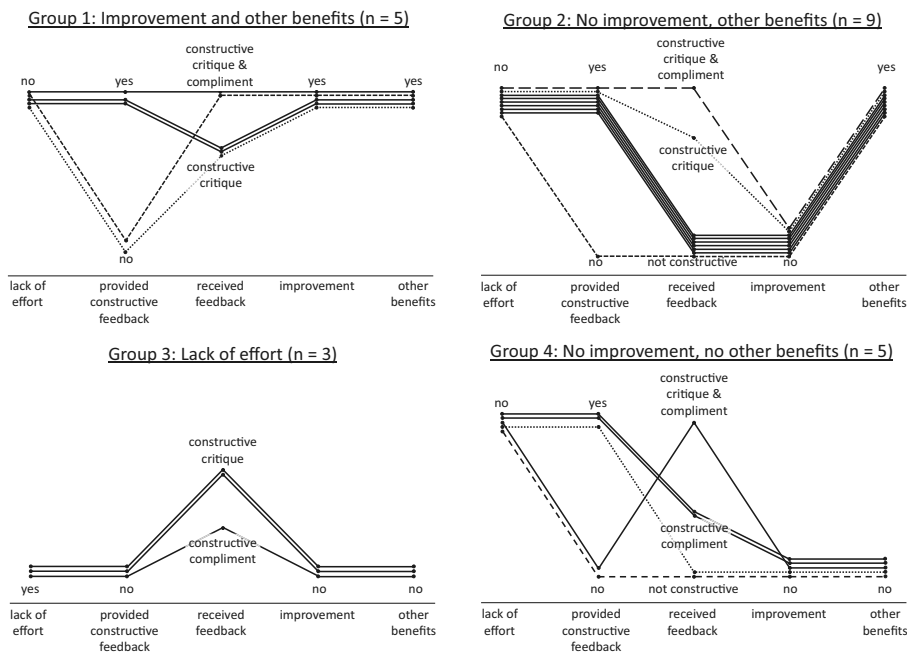


Figure 2 Students' PA pathways. Each line represents one student and his/her experience of this PA

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)

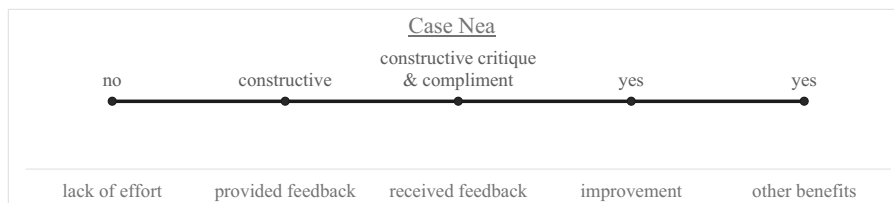


Figure 3 Nea’s pathway through PA




Measurements are reasonable and clearly presented.				<i>Measurements were clearly in the text and seemed reasonable</i>
----------------------------------------------------	-----------------------------------------------------------------------------------	-----------------------------------------------------------------------------------	-----------------------------------------------------------------------------------	--------------------------------------------------------------------

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.				<i>You could have added the equation to calculation</i>
-------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	---------------------------------------------------------

Figure 5 Example of constructive critique Nea provided




Research plan and equipment have been clearly presented.				<i>You could have told more about materials that you used and how you made the measuring.</i>
----------------------------------------------------------	-----------------------------------------------------------------------------------	-----------------------------------------------------------------------------------	-----------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------

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

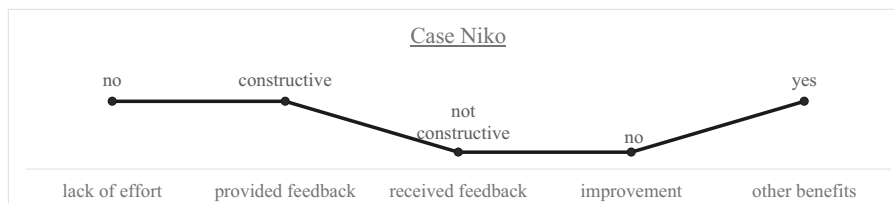


Figure 7 Niko's pathway through PA




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

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.				<i>These are quite clearly written!</i>
--------------------------------------------------------	-------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------	-----------------------------------------

Figure 9 Example of unconstructive compliment Niko received

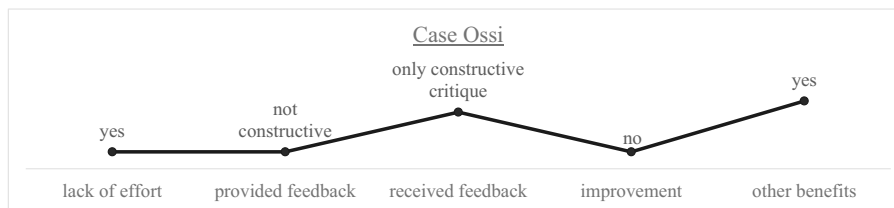


Figure 10 Ossi's pathway through PA

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

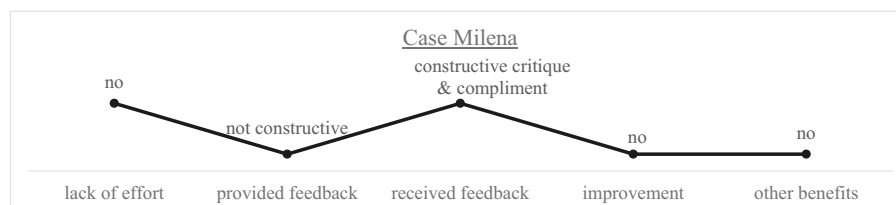


Figure 11 Milena's pathway through PA

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.

Funding Information Open access funding provided by University of Jyväskylä (JYU).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anker-Hansen, J., & Andréé, 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]. Helsinki, Finland: Opetushallitus.
- Finnish National Board of Education (FNBE). (2014). *Perusopetuksen opetussuunnitelman perusteet 2014* [National Core Curriculum for Basic Education 2014]. Helsinki, 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.



II

DEVELOPMENT OF SECONDARY STUDENTS' FEEDBACK LITERACY: PEER ASSESSMENT AS AN INTERVENTION

by

Laura Ketonen, Pasi Nieminen & Markus Häikiöniemi, 2020

The Journal of Educational Research, 113(6), 407–417.

<https://doi.org/10.1080/00220671.2020.1835794>

Reproduced with kind permission by Routledge.

The development of secondary students' feedback literacy: Peer assessment as an intervention

Laura Ketonen , Pasi Nieminen , and Markus Häikiöniemi 

Department of Teacher Education, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

A growing body of research has recognized the importance of students' having active roles in feedback processes. Feedback literacy refers to students' understandings of and participation in feedback processes, and research on students' feedback literacy has so far focused on higher education; secondary schools have not received attention. This case study investigates secondary students' feedback literacy and its development in the context of formative peer assessment. From various data sources, three categories of students' feedback literacy were identified, and criteria for the levels of literacy in each category were created. The criteria were used in the coding of seventh- and eighth-grade students' skills. The results show that students were able to develop their feedback literacy skills. Thus, secondary school students should be introduced to feedback literacy via, for example, formative peer assessment.

ARTICLE HISTORY

Received 18 May 2020
Revised 18 August 2020
Accepted 7 October 2020

KEYWORDS

Feedback literacy; peer assessment; formative assessment; feedback; secondary education; science

Introduction

Feedback can significantly enhance or inhibit learning (Hattie & Timperley, 2007), and it has therefore generated great interest and considerable research. Ideally, feedback helps students achieve their goals (Hattie & Timperley, 2007). The studies on feedback have mainly concentrated on the features of efficient feedback, thus implying that the responsibility of a successful feedback process lies on the provider of feedback, who must consider its content, tone, and timing. Similar attention to the quality of feedback can be seen in research on peer assessment. Only recently has receiving feedback generated steady attention from scholars (Boud & Molloy, 2013; Delva et al., 2013; Jonsson, 2013; McLean et al., 2015; Sutton, 2012; Wiliam, 2012; Winstone et al., 2017). Carless and Boud (2018) introduced a framework of feedback literacy describing the competences necessary to participate in feedback processes; it has generated considerable interest and encouraged further research. Studies have investigated students of higher education (Han & Xu, 2019a, 2019b; Hey-Cunningham et al., 2020; Molloy et al., 2019) and academics (Gravett et al., 2020), but so far no study has investigated the feedback literacy of a much larger group: secondary students. Exploring the feedback literacy of secondary school students is the first goal of this study.



Carless and Boud (2018) expressed the need for research on the development of students' feedback literacy combined with intervention. This article is such a study, and its second goal is to investigate the development of lower secondary school students' feedback literacy, when formative peer


assessment is repeatedly used in students' physics and chemistry lessons. Peer assessment is used not only as a tool to uncover students' perceptions about assessment and feedback, but also as an intervention that provides students with the opportunities to reflect and practice feedback processes and hence to develop their feedback literacy.

Feedback literacy

Sutton (2012) conceptualized feedback literacy as “the ability to read, interpret and use written feedback” (p. 31). Carless and Boud (2018) built on his work and defined feedback literacy as “the understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies” (p. 1316). Feedback literacy highlights the need for students' own activity in feedback processes. Though teachers provide feedback, no matter how useful it is, it does not automatically benefit the receiver. The feedback needs to be accepted, processed, and acted on by the receiver. Feedback literacy denotes these competences.

The framework of Carless and Boud (2018) presents four features of students' feedback literacy, which are appreciating feedback, making judgments, managing affect, and taking action. Appreciating feedback includes understanding that feedback is for improvement, understanding that the recipient must have an active role in the feedback process, appreciating different forms and sources of feedback, and the ability to use technology in feedback processes. Making

CONTACT Laura Ketonen  laura.k.ketonen@jyu.fi  Department of Teacher Education, University of Jyväskylä, PO Box 35, FI-40014 University of Jyväskylä, Jyväskylä, Finland.

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/00220671.2020.1835794>.

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

judgments comprises capacities of judging one's own and others' work, participating productively in peer feedback processes and developing self-evaluative skills. Managing affect includes maintaining emotional balance, avoiding defensiveness, having dialogue about feedback, and striving for continuous improvement. These three features (appreciating feedback, making judgments, and managing affect) add students' possibilities for the fourth feature, which is taking action. Taking action comprises understanding that using feedback requires the recipient's activity and developing strategies for acting on feedback.

The work of Carless and Boud (2018) has induced further research. Molloy et al. (2019) explored higher education students' feedback literacy from the students' perspective. Their results present seven groups containing altogether 31 categories of knowledge, capabilities, and skills of feedback literacy. Hey-Cunningham et al. (2020) designed a pilot of blended learning for postgraduate research students in academic writing that contained principles, exemplars, self-assessment, and peer assessment, and noticed that it improved students' feedback literacy. In particular, students felt that they learned skills and gained experience about how to proceed with the received feedback. Han and Xu (2019b) explored the development of higher education students' feedback literacy, and used teachers' feedback on peer feedback as an intervention. Researchers reported a growth in feedback literacy, especially in students' ability to assess their peers' work, that is, judging the quality and providing feedback for learning. They reported a considerable individual variation in students' development. Han and Xu (2019a) investigated higher education students' feedback literacy and its impact on engagement in the context of teachers' written corrective feedback. They found that individual students' skills of feedback literacy were unequally developed, which limited students' engagement with feedback. The researchers noticed not only that students' feedback literacy was dynamic and developed in the feedback processes, but also that the development appeared rather randomly.

Peer assessment

Building on previous research (Carless & Boud, 2018; Sadler, 1989; Topping, 2013), formative peer assessment is defined as a process in which students evaluate or are evaluated by their peers with the intention that both the assesses and assessee enhance their work or learning strategies in the process. When discussing peer assessment, the concept of feedback must be understood broadly, as both assessing and being assessed provide opportunities for receiving feedback. Carless and Boud defined feedback as "a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies" (p. 1315). In peer assessment, the potential sources of information are many, including assessment criteria, assessed work, received feedback, interaction with classmates, and coaching from the teacher.

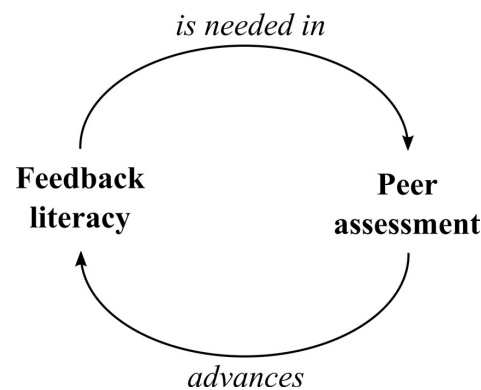


Figure 1. Relationship of feedback literacy and peer assessment.

Relationship of peer assessment and feedback literacy

Peer assessment and feedback literacy are inter-related (Figure 1). Participating productively in peer assessment requires feedback literacy (Han & Xu, 2019b), and peer assessment provides opportunities to advance it. In the best case, feedback literacy and peer assessment support each other, as positive experiences of peer assessment predict teachers' frequent use of peer assessment (Panadero & Brown, 2017). Students' advancement in feedback literacy makes peer assessment a more applicable tool; therefore, feedback-literate students are more likely to get further opportunities to develop their feedback literacy with peer assessment. Researchers have a common understanding that peer assessment requires training (Gielen et al., 2010; Hovardas et al., 2014; Lu & Law, 2012; Topping, 2009; van Zundert et al., 2010), which implies that a certain level of feedback literacy is needed to enable a productive peer assessment process. If students are not feedback-literate enough to benefit from peer assessment, the practice is more likely experienced as malfunctioning and rejected.

The prior research on peer assessment does not generally mention feedback literacy but recognizes its features. The next sections introduce research that presents how each of four features of feedback literacy (Carless & Boud, 2018) are needed in peer assessment as well as research that shows that peer assessment has the potential to advance each feature of feedback literacy.

Appreciating feedback

When students understand that feedback is a learning tool, they are more likely to benefit from peer assessment (Ketonen et al., 2020). Incomprehension in this area may result in unproductive practices, such as the provision of superficial positive feedback (Tasker & Herrenkohl, 2016) or friendship marking (Foley, 2013; Panadero et al., 2013), which occurs when students let social relationships affect the feedback they provide. Peer assessment can help tackle these issues because it is a tool for discussing the different aims of summative and formative assessment (Davis et al., 2007). Tasker and Herrenkohl (2016) reported that peer assessment provides a context for training and reflection that advances

students' appreciation of critical feedback and leads them from providing positive, superficial peer feedback to providing guidance. Another essential component of feedback literacy in peer assessment is appreciating feedback from different sources. Students tend to undervalue and disregard feedback from their peers (Foley, 2013). Peer assessment can help with this issue as well because it advances students' appreciation of peers as a source of feedback (Crane & Winterbottom, 2008).

Making judgments

Peer assessment requires the skills of judging received feedback and judging others' work. Concerning judgments, Carless and Boud's (2018) framework for feedback literacy concentrates on judging one's own and others' work, and Molloy et al. (2019) further included the aspect of processing received feedback. The ability to critically interpret feedback is an important skill that enhances the benefits of peer assessment (To & Panadero, 2019). Guiding students to be active in the assessee's role by evaluating received feedback improves their motivation to engagement in peer assessment (Minjeong, 2009). Students need to understand the assessment criteria in order to judge their peers' work (Cartney, 2010; Foley, 2013; Panadero et al., 2013). Moreover, a shared understanding of the criteria increases students' comfort with peer assessment (Panadero et al., 2013). Peer assessment can be used to train students to deepen their understanding of the criteria (Anker-Hansen & André, 2019; Black & William, 2018) and develop the ability to judge others' work (Han & Xu, 2019b).

Managing affect

Peer feedback may raise negative emotions (Cartney, 2010; Panadero, 2016). Assesseees can be defensive against corrective peer feedback (Anker-Hansen & André, 2019; Ketonen et al., 2020; Tasker & Herrenkohl, 2016), and assessors may worry about assesseees' negative emotions (Cartney, 2010; Davis et al., 2007). However, because of peer assessment's interactive nature, affective issues cannot be avoided (Panadero, 2016). Managing affective issues is hence necessary, and for productive peer assessment, students need support with the emotional aspects related to both the assessors' and assesseees' roles (Cartney, 2010). Fortunately, peer assessment improves psychological safety within a group (van Gennip et al., 2010), meaning that group members feel safer sharing their opinions and asking for help. This promotes elicitation of suggestions from peers and discussions about them, which relate to managing affect in Carless and Boud's (2018) framework.

Taking action

Studies have noted students' reluctance to revise their work according to peer feedback (Anker-Hansen & André, 2019; Tsivitanidou et al., 2011; 2012). As theorized by Carless and Boud (2018), this reluctance is the result of insufficient feedback literacy, such as an insufficient understanding of assessment criteria (Tsivitanidou et al., 2011) or formative assessment (Ketonen et al., 2020). As explained previously,

peer assessment can help address these issues and is generally considered useful in encouraging students to act on feedback (Jonsson, 2013), especially when integrated in a deeper conversation about assessment (Cartney, 2010).

Research questions

Feedback literacy seems an inherent requirement of productive peer assessment, but it is little researched at the secondary level. In order to investigate secondary students' feedback literacy and its development, the following research questions were addressed:

Research Question 1: What kind of skills of feedback literacy do students have in the context of formative peer assessment?

Research Question 2: How does students' feedback literacy develop during one year of using peer assessment repeatedly?

Method

Participants

The study was carried out in an ordinary urban school in Finland. The intervention was conducted in cooperation with an experienced subject teacher and her two science classes (15 and 16 students, respectively); the study started at the beginning of the students' Grade 7 year and lasted until the middle of their Grade 8 year. At the beginning of the intervention, the students did not know each other well, but most had at least one friend in the class. In Finland, students study general science in Grades 1–6 with a class teacher and begin physics and chemistry in Grade 7 with a subject teacher. This transition into more specialized learning seemed like an ideal moment to introduce a new practice. Peer assessment is not yet a well-established practice in Finland (Atjonen et al., 2019), and when asked, approximately half the students had experiences with peer assessment, though none had used it regularly.

The teacher was asked to join the study because of her well-organized, thoughtful working style, which enabled the intervention's careful co-planning, and for her considerate but not overly individual teaching style. The study aimed at understanding and explaining the phenomena of peer assessment and feedback literacy instead of making broad generalizations, and the number of participants was kept small in order to enable knowing the students and gathering of rich data. Two students chose not to participate in the study; the total number of participants was 29. Twenty-two students participated in all parts of the first peer assessment, and 15 of those 22 participated in all parts of the second peer assessment, which allowed the researcher to track their development. Two of 15 students missed three or more of the five peer assessments that were arranged between the abovementioned two peer assessments, which diluted the intervention, and they were therefore left out of the study. Altogether, 13 students provided a satisfying dataset, and they were included in the analysis of the study.




	Lacking 	Partly done 	Everything is ok 	Comment What was good? What could be improved and how?
The equipment is listed clearly. Another group could repeat the inquiry with the information.				
The research plan is explained clearly. It can be understood without already knowing the instructions.				
The results show the value of the current of several wire lengths.				
The results include a graph showing how the length of the wire affects the current.				
The axes are named, equally spaced, and have units.				
Error of measurements and results are carefully explained.				

Figure 2. Premade assessment criteria of PA2.

Procedure

The study examined students in two similar peer assessments. The first (PA1) was conducted during the fall semester of Grade 7, and the second (PA2) was conducted a year later. In PA1 and PA2, the students conducted and reported on a physics inquiry. The topics (determining the speed of an object and determining the resistance of a wire) were different due to the curricula, but the difficulty of the tasks, the form of the lab reports, and the assessment criteria were similar. The students conducted the inquiry in groups but created the reports individually. After the reports were finished, they were assessed by their peers. The researcher and the teacher planned the pairings so that students were not partnered with someone they conducted the inquiry with and with sensitivity to social issues, such as pairing vulnerable students with considerate classmates. The students were provided with premade criteria in which each criterion was assessed with a 3-point Likert scale and optional written comments (Figure 2). After the peer assessment, the reports and feedback were returned to the students, and they had time to improve their work before returning it to the teacher for summative assessment.

In addition to these two peer assessments, training sessions and several other peer assessments were arranged. Before PA1, the students had six training sessions, and they practiced peer assessment once. Between PA1 and PA2, they had five more peer assessments and one training session. All parts of the interventions are presented in Table 1. The training sessions and peer assessments were embedded in the curricula and consisted of class discussions led by the

researcher or the teacher, both of whom used cards displaying examples of feedback (Figure 3), principles of peer assessment, written tasks (e.g., creating or testing assessment criteria), and actual peer assessments. The training focused on issues that previous research noted as important for successful peer assessment, specifically understanding the idea of formative assessment, understanding of the qualities of useful feedback, internalizing assessment criteria, comparing the criteria with the work, social rules, and practicing actual peer assessment. The key message was that peer assessment is for helping others learn and receiving help from others.

The students remained focused during the class discussions and seemed to find the topics interesting. The intention of the discussions was not to find “the right answer,” but rather to encourage the students to share their thoughts. The training was flexible, as the intent was to respond to the students’ needs. For example, when peer assessment was first practiced, the students struggled to deal with critical feedback; the issue was therefore included in the training and discussed before PA1. The teacher and the researcher had prolonged experience in co-teaching at the school, which helped co-planning the interventions and made sharing the responsibility in the classroom natural. During the written tasks and peer assessments, they circulated throughout the classroom and held discussions with students who needed help, encouraging them to use and trust their own judgment. Even though class sessions were planned so that the students were equipped with the necessary information, discussions with individual students and working groups appeared necessary and influential.

Table 1. Interventions of the study.

Time	Activity	Description
Grade 7 fall	Six sessions of training peer assessment skills (10–45 min)	<ul style="list-style-type: none"> • Reflecting on the use of assessment • Reflecting on the quality of feedback • Creating assessment criteria • Comparing a work to created criteria • Self-assessment with created criteria • Reacting to feedback
	Peer assessment (task 40 min, assessment 30 min, improvement 10 min)	Groups assessed other groups' inquiry plans. Afterwards, groups had time to improve their work.
	PA1 (task 3 hr, assessment 45 min, improvement 45 min)	Each student assessed another student's lab report with provided criteria. The researcher and the teacher planned the pairings. Afterwards students had time to improve their work.
Grade 7 spring	Interview	Discussing PA1
	Five peer assessments (approximately 5 min)	Working pairs provided each other with feedback based on criteria after four peer assessments.
	Peer assessment (task 15 min, assessment 10 min, improvement 10 min)	Each student assessed another student's inquiry plan with provided criteria. The researcher and the teacher planned the pairings. Afterwards students had time to improve their work.
Grade 8 fall	Interview	Discussing all previous PA sessions.
	Peer assessment training (30 min)	Students categorized the feedback comments that they provided during PA1. Intervention was adapted from Tasker and Herrenkohl (2016).
	PA2 (task 3 hr, assessment 30 min, improvement 30 min)	Each student assessed another student's lab report with provided criteria. The researcher and the teacher planned the pairings. Afterwards students had time to improve their work.
	Interview	Discussing PA2

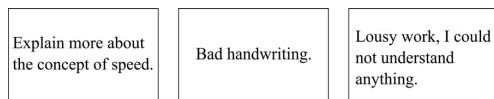


Figure 3. The feedback example cards discussed in one training session. The class agreed that the first critique was worth considering, whereas the third was unhelpful, offensive, and not worth keeping.

Research design and data

A case study design was adopted for the exploratory nature of the research. The case study approach allowed us to know all the students and gather versatile data. The data include field notes and audio recordings of lessons, student interviews, and students' written work and feedback. The researcher placed recorders on the tables of each student pair and took field notes on each lesson in which she participated (approximately 60 lessons of 1.5 hr each). Students' original and revised work and provided feedback were scanned. Besides the analysis, these documents were used as a basis of conversation in semi-structured, audio-recorded student interviews soon after PA1 and PA2.

Analysis

The analysis comprised three parts. In the first and second parts, the data were analyzed using a thematic analysis (Braun & Clarke, 2006) to identify and report patterns in the data. A thematic analysis was chosen for its flexibility: it is not bound to any specific theoretical framework, and it works well in analyzing data consisting of various forms. It can be used to find similarities and differences within the data, and it can produce unexpected insights. The patterns can be driven by either data or theory.

The first part of the analysis was driven by theory. The aim was to recognize the features of feedback literacy that appeared in the data during peer assessment and to adjust

the features of Carless and Boud's (2018) framework to the context of formative peer assessment. We looked for data extracts that contained information on the students' feedback literacy and attached codes to them. The codes described the students' actions and attitudes relating to the elements of feedback literacy: appreciating feedback, making judgments, managing affect, and acting on feedback. After coding the whole dataset, the codes were gathered into preliminary categories. The distinctiveness of the categories was then examined, and the categories were adjusted and named. The data were recoded, and categories were re-examined and readjusted until no further changes were undertaken. In the end, three categories of feedback literacy skills were discerned.

The goal of the second part of the analysis was to examine the skills that students had in the three categories of feedback literacy. During the first part of the analysis, we noticed students had varying feedback literacy skills, and we defined and described them in an iterative process. The analysis was driven by data, but its scope was limited to the previously identified categories. Within each category, we formed case groups with similar skill levels. Then, we examined and described the groups and returned to the data with revised descriptions. We revised the groups until new rounds did not produce changes to the groups or their descriptions. While remaining sensitive to the theory, we organized the groups from the most basic to the most advanced, thereby creating a criteria-based rubric for feedback literacy skills. In two cases, a skill level was created from the data of only one student, and we enriched that data with interviews with additional students in the spring of Grade 7 (Table 1). This enabled us to test our rubric and find more examples to describe the levels.

In the third phase of analysis, the category criteria were used in evaluating seventh- and eighth-grade students' feedback literacy, and multiple data sources were also used. We coded students' "Understanding of the purpose of feedback"

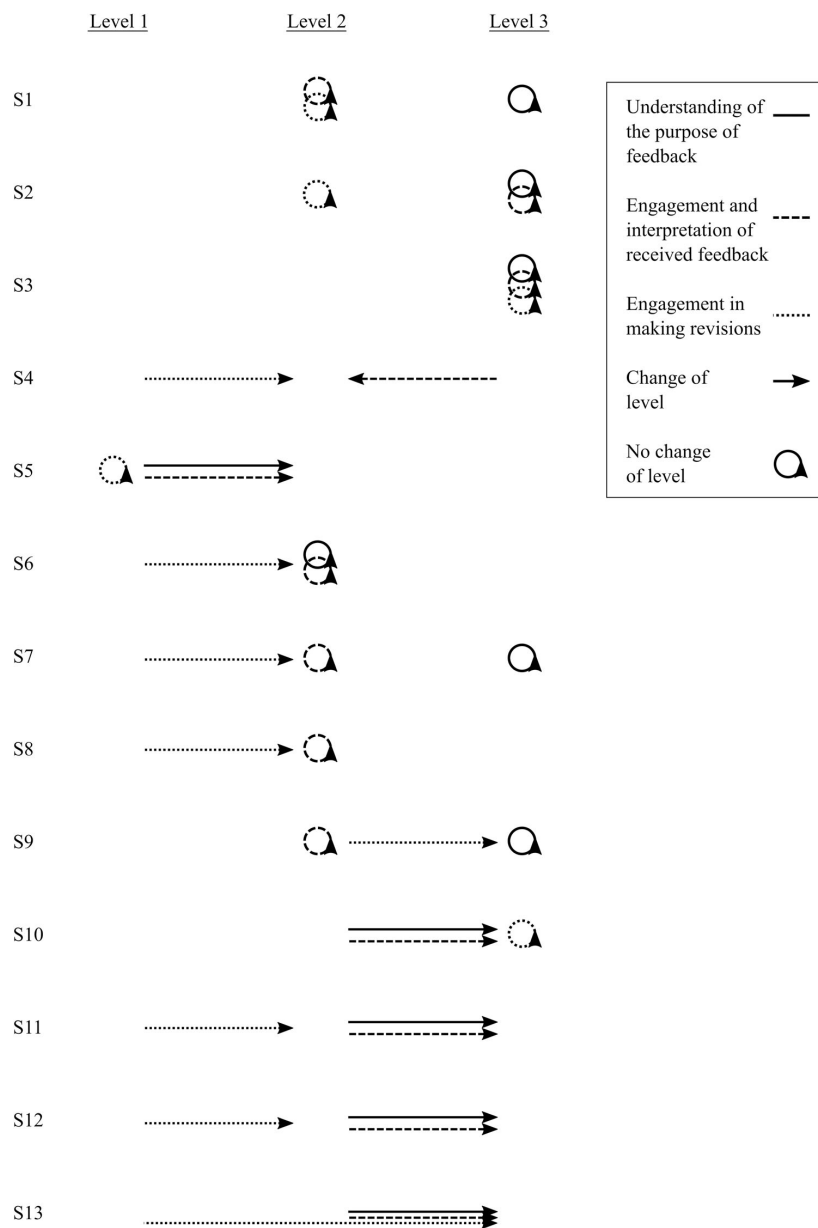


Figure 4. The change in students' feedback literacy.

and “Engagement and interpretation of received feedback” skills using interviews as a primary data source, and to ensure reliability, we compared students’ narratives to their own work, their received and provided feedback, and the work that they had assessed. If the findings needed further affirmation, we used field notes from and audio recordings of the lessons. In two cases (Students 4 and 8), even these data did not provide enough information to determine the category level, and these skills were left out of these students’ results (Figure 4). The “Engagement in making revisions” skill was coded using students’ original and revised

work to recognize the changes they had made. Their engagement in making these changes was interpreted using interviews with students, field notes from the lessons, and audio recordings of the lessons.

This study’s authors engaged in peer negotiation to test the levels and categories. Coding students’ skills was a complex task that required precise interpretation and the management of versatile data because students expressed their feedback literacy skills in diverse parts of the interview, and as explained previously, the comments were interpreted with other data. Therefore, negotiation among the authors was

Table 2. Understanding of the purpose of feedback.

	Label	Criteria	Examples
Level 1	No appreciation of feedback	Student expresses that he/she does not appreciate feedback.	R: Could the feedback [have] been such that you would have made some changes [to your work]? S: No ... R: What do you think, if the teacher had given you feedback, would it have been more effective or the same? S: Exactly the same. (Student 5, PA1)
Level 2	Appreciation of positive feedback	Student expresses that receiving positive feedback is a satisfying or hoped-for situation.	R: Would you say that you benefited from the feedback? S: Yes. R: How? S: It was positive, like it made me feel good and then you know that you did well. (Student 6, PA1)
Level 3	Appreciation of corrective feedback	Student expresses that receiving corrective feedback is a hoped-for situation.	R: What did you think of having positive feedback? S: Well, that did not help me much ... if I, like, I was not satisfied with this last part [of my work] because I ran out of time ... so I hoped for help with what to add there. (Student 11, PA2)

Table 3. Engagement and interpretation of received feedback.

	Label	Criteria	Examples
Level 1	No interest in feedback	Student expresses that he/she is not interested in engaging with the feedback.	R: Okay, you received the feedback. I have it here. Erm, so what did you think when you received the feedback? S: Nothing. R: Okay. Did you read it? S: Nope. R: Okay. So you did not think anything. Why didn't you read [it]? S: Since I am not, like, interested what it says. (Student 5, PA for seventh grade spring)
Level 2	Reading feedback	Student expresses that he/she has read the feedback but does not show signals of active interpretation of feedback compared with his/her own work OR has problems with selectively rejecting the data.	R: Tell me what you did. S9: So I added there that it [speed] is distance divided by time. Then I wrote that more accurately, the calculation. Since it is there [in the original work], a little small, so ... <i>When receiving feedback, the student found the comment of missing calculation amusing, since it was clearly written on the report. Nevertheless, the student did not reject the comment but wrote the same calculation in bigger letters.</i> (Student 9, PA1)
Level 3	Active interpretation of feedback	The data show that the student actively interprets the feedback by comparing it to his/her own work and, if he/she finds it necessary, rejects the feedback.	S12: I took, like, at least the most important ones [feedback comments] and then ... R: Yes? S12: Well, the others I ignored, there was not time for everything. (Student 12, PA2)

considered appropriate to examine the reliability of coding. Laura Ketonen, who was the only one involved in teaching, prepared the data of three regular and two specifically challenging cases for negotiation. The data from each student were gathered into one document, and the data extracts used in defining the levels of the categories were discussed. Pasi Nieminen and Markus Hähkiöniemi explained their views on the Laura Ketonen's coding, and discrepancies were carefully discussed. After the regular cases, the challenging ones were examined. We had a mutual agreement that, due to scant interviewee input from students, these cases were considerably more challenging than the regular cases, and definitively determining the level of students' skills was not possible. The decision to leave one skill out of the results in two cases (Students 4 and 8) was made in the meeting.

Results

The results are twofold. The first part of the results consists of the categories with levels of students' feedback literacy

(Tables 2, 3, and 4), and the second part consists of the paths of development of secondary students' feedback literacy.

Categories and levels for students' feedback literacy

Three categories of students' skills of feedback literacy were recognized in the context of formative peer assessment: understanding of the purpose of feedback, engagement and interpretation of received feedback, and engagement in making revisions. Understanding of the purpose of feedback (Table 2) relates to students' dispositions to feedback. The students on the first level do not appreciate any kind of feedback. The second level, appreciation of positive feedback, reflects the understanding of feedback as a judgment. By this conception, corrective feedback communicates a shortcoming or a failure, and it is not wanted. Conversely, students on the third level welcome corrective feedback, as they see it as a mediator of learning and improvement.

Engagement and interpretation of received feedback (Table 3) describes students' activity in judging the received feedback. The first level is not being interested in feedback

Table 4. Engagement in making revisions.

	Label	Criteria	Examples
Level 1	No improvement	Student does not improve his/her work.	R: ... Did you make any changes? S: Erm, I don't think I made, yes, I think it [the revised work] is exactly the same [as the original]. (Student 7, PA1) <i>There was a need for improvement in the student's report, but he made no changes.</i>
Level 2	Superficial improvement	Student makes improvements that do not require much activity nor learning new knowledge or skills.	R: Would you say that it [the feedback] was useful? S: Yes, I fixed them. <i>There was a significant need for improvement, but the student quickly added two missing numbers and one word to his report.</i> (Student 1, PA1)
Level 3	Investment in improvement	Student improves work as a result of his/her own activity, for example, seeking new feedback, guidance or new knowledge, or practicing new skills.	S: Yes, we started to wonder in research group, since we all had almost [a comment like] that, like all the others had similar results, and we started to think what we had done wrong. ... So it was a huge effort to figure it out. <i>Student remade the measurements and asked for extra time to complete the work at home.</i> (Student 13, PA2)

at all, and it recalls the first level of the previous category. This is natural, as students' passive role and reluctance to accept feedback is reflected in several aspects of feedback literacy. The second level describes the student reading the feedback but staying in the passive role of a receiver. The third level comprises active interpretation and evaluation of the received feedback.

Engagement in making revisions (Table 4) describes students' active engagement in making revisions. Do students improve their work, and if so, do they merely follow the instructions of the feedback, or do they have an active role in finding new knowledge or feedback, for example from the teacher, the material or their peers?

Development of students' feedback literacy skills

The development of students' feedback literacy in Grades 7 and 8 is presented in Figure 4. Different line types represent the different categories of feedback literacy. Arrows show the changes in students' feedback literacy from Grades 7 to 8, and circles indicate staying on the same level. Changes primarily happened from the lower to the higher level. Improvement was noticed in all categories: understanding of the purpose of feedback (five cases), engagement and interpretation of received feedback (four cases), and engagement in making revisions (seven cases). One transition happened from the higher to the lower level in engagement and interpretation of received feedback.

Next, we will present students' development of feedback literacy with two examples. The first example is used to demonstrate improvement in the understanding of the purpose of feedback and the second students' engagement and interpretation of received feedback.

Example: Improving the understanding of the purpose of feedback

Student 10 showed strong development in feedback literacy. She was already on the highest of three levels on "engagement in making revisions" in PA1, but in "understanding of the purpose of feedback" and "engagement and interpretation of received feedback," she

advanced from Level 2 to Level 3. In PA1, the researcher inquired into the student's feelings about the received, predominantly positive feedback:

R: When you read the feedback, how did you feel about it?

S10: It was good, like I had thought about the task quite a lot and tried to make it, like, as good as possible, so it is encouraging when you kind of succeed.

R: Was the feedback useful to you in some way?

S10: Yeah. Or it was fun, like, have your classmate assess you. That was sort of useful. I don't know. It was nice anyway.

The student stated that she had enjoyed the positive feedback. She could not actually say that it was useful, but nevertheless, she was contented with receiving positive feedback. A year later, the researcher asked student the same question about PA2:

R: You received this feedback. So what did you think about it?

S10: Like, it was pretty good, like I received positive feedback, but maybe, like, it could have been, like, I don't know, cause there was nothing that I could have improved or anything.

R: Would you have hoped for something like that?

S10: Well, maybe.

R: Mm?

S10: Yes. So I could have used it for improvement.

A year later, the student was no longer happy with positive feedback. She stated that she had hoped for guiding feedback that would have helped her to improve her work. Her expectations for feedback had changed as she had learned to appreciate corrective feedback.

The student's development in "engagement in making revisions" was connected to learning to appreciate critical feedback. In PA1, the student only made the one change that was suggested in the feedback. In PA2, she actively looked for feedback. Even though her peer assessor did not provide guidance, she utilized the feedback that her work partner had received and the feedback that assessing the other student's work provided her. She had adopted a view of feedback as being meant for improvement and become more active in making revisions.

Example: Improving engagement and interpretation of received feedback

Student 12 was less feedback-literate than Student 7 in the earlier example. In PA1, Student 12 was on Level 1 on “engagement in making revision” and Level 2 on “understanding of the purpose of feedback” and “engagement and interpretation of received feedback.” This meant that he superficially performed his role as a peer assessor and did not revise his work. After PA1, the researcher asked the student’s thoughts about the received feedback:

R: Was it useful for you, the feedback?

S12: Yes, it was.

R: Tell me how?

S12: Well, there is, for example, that I could tell more clearly how to make the measurements ... there was, for example, that it [conducting measurement] was told clearly, so it points out how to make it in the future.

The student’s narration of the feedback is contradictory. He had received a feedback comment that the work was hard to interpret. Student 12 had looked at the feedback but not worked to understand its meaning. Instead of comparing the feedback to his work and deciding to reject the feedback or to improve the work, he explained that he had taken the feedback as general guidance for the future (Level 2). A year later, after PA2, the student explained his interpretations of his feedback as follows:

S12: I had thought that if I received good tips, I could improve my work a little, since I hadn’t come up with ideas myself. But then [after receiving feedback] I started figuring out what I should do and, erm, like, I was running out of time, so I finished this [shows a part of the task]. And, erm, I added here ...

R: Yes?

S12: I took, like, at least the most important ones [feedback comments] and then ...

R: Yes?

S12: Well, the others I ignored, there was not time [for] everything.

The student explained that he had evaluated which were the most important feedback comments and ignored less important ones. In addition, he stated that he had considered part of the feedback comments wrong and therefore ignored them (Level 3). When the researcher inquired about an example of such a comment, he answered:

S: Well, maybe this one, since it said, “how many wires”, since, erm, usually you can take enough and, erm, like, see how many you will use.

The student’s engagement and active interpretation of received feedback advanced from PA1 to PA2. In PA2, the student had several strategies for dealing with feedback: to approve, to prioritize, and to reject. The student did not improve his work after PA1, but after PA2, he made two meaningful improvements that he had judged important to make. Active interpretation of received feedback appeared to enable him to revise his work meaningfully.

Discussion

The present study explored the secondary students’ use of formative peer assessment and showed that seventh-grade students developed feedback literacy skills after one year of practicing peer assessment. The research on feedback literacy has so far focused on higher education, but these results demonstrate that Carless and Boud’s (2018) framework is relevant at the secondary level. Prior studies did not explicitly describe the levels of students’ feedback literacy (Han & Xu, 2019b; Hey-Cunningham et al., 2020), but this study identified three levels of development in three categories of feedback literacy. Building an understanding of the levels of development in feedback literacy will enable the tracking of students’ longitudinal development across disciplines and school sectors, and it will help educators identify students’ strengths and needs. In this study, the levels were used to determine seventh- and eighth-grade students’ feedback literacy. On average, advancement was noticed in all three categories throughout the year of practicing peer assessment in their science studies, but individual variations were noticed, as they were in the context of higher education (Han & Xu, 2019b). Han and Xu (2019b) and Hey-Cunningham et al. (2020) explored and reported development in graduate and postgraduate students’ feedback literacy, respectively. Features in common with the present study were the explicit discussion of the feedback processes and the use of peer feedback, implying that these are promising practices to promote the development of students’ feedback literacy.

The identified categories of feedback literacy in which the students showed different skill levels during peer assessment were “understanding the purpose of feedback,” “engagement and interpretation of received feedback,” and “engagement in making revisions.” The levels of the first category—“understanding of the purpose of feedback” (Table 2)—described the students’ expectations of feedback: did they want it, and if so, did they want to be told they had done a good job or find opportunities for further improvement? The students’ development was shown through their increased interest in feedback and appreciation of criticism. Other studies have presented similar findings. Tasker and Herrenkohl (2016) found that students learn to value critical feedback, and Han and Xu (2019b) reported a cultivated willingness in students’ to ask for peers’ opinions. Assessment practices in Finland are prevalently summative (Atjonen et al., 2019), and students are accustomed to receiving evaluative feedback at the end of courses when criticism cannot be used to improve their performance. In this study, peer assessment was used to emphasize formative assessment. Peer feedback was specifically designated for improvement, and the time given for revising work strengthened this message. In formative peer assessment, students are persuaded to focus on improvement and hence gain experience with corrective feedback and find it desirable.

The levels of the second category—“engagement and interpretation of received feedback” (Table 3)—related to the students’ actions when receiving peer feedback: did they read the feedback, and if so, were they active in interpreting it and able to reject it when reasonable? Hey-Cunningham

et al. (2020) noticed a similar scale of conception with post-graduate students. A considerable portion of students in their study indicated that they would not question supervisors' feedback, while some expressed that they would think about it critically. After attending a program concerning the use of feedback, all students understood the value of critically considering feedback. In the present study, the critical reception of feedback appeared challenging to the students. They tended to expect clear, unambiguous feedback that did not require active interpretation, a factor also found by O'Donovan (2017) and one that stems from becoming accustomed to teacher feedback that is not expected to be questioned or rejected. Receiving feedback from a peer exposed the students to the need to judge it. Additionally, the topic was explicitly addressed, and the appraisal of feedback was encouraged.

The levels of the third category—"engagement in making revisions" (Table 4)—described whether the students improved their work and, if so, whether that was due to their following the explicit advice of the feedback or as a result of their own activity. Hey-Cunningham et al. (2020) found similar levels of feedback literacy, stating that post-graduate students learned to understand revision as a process that was not about following feedback but engaging in learning. Carless and Boud (2018) theorized that the ability to take action stems from developing other feedback literacy skills. Understanding corrective feedback as a mediator of improvement is pivotal in peer assessment (Ketonen et al., 2020; McConlogue, 2015; Tasker & Herrenkohl, 2016), and the results of this study support that idea. Several studies on secondary students mentioned their struggles to act on peer feedback (Anker-Hansen & Andrée, 2019; Hovardas et al., 2014; Tseng & Tsai, 2007), and we suggest that the reason might be students' undeveloped feedback literacy skills.


This study has implications for practice. The levels of the first category show that the students were not necessarily inclined to read feedback or use it for improvement. In such cases, there is no point in pushing feedback, but the issues concerning students' goals and their understanding of feedback must be dealt with. Feedback cannot help students achieve their goals if they do not have any or if students construe guiding feedback as offensive criticism. The levels of the second category revealed the students' passive approach to feedback. Even when interested in it, most of the students were not initially prepared to invest in interpreting it but expected the feedback to tell them unambiguously how to proceed. If students are not accustomed to judging feedback, they need to be encouraged to do so. The freedom of whether to act on feedback may assist in seeing critical feedback as a gift that potentially advances learning instead of a burden that automatically leads to more work.


We acknowledge several study limitations. The first is that the research is not experimental; therefore, it is unclear to what extent the use of peer assessment affected the development of students' feedback literacy. Physics and chemistry contexts comprise only 5% of studies in lower secondary schools, and during the school year, the students naturally faced experiences other than this intervention. The second


limitation is this study's specific context, as it focused on two student groups at the same school. The third limitation relates to the method. Methodologically, this study differs from previous ones exploring the development of feedback literacy (Han & Xu, 2019b; Hey-Cunningham et al., 2020) because it created criteria for the levels of feedback literacy and thus made the features of development explicit. However, other feedback literacy categories could probably be found with more data sources. For example, the managing of affect could not be detected, as the students were predominantly private about their negative feelings. Hence, further research on the variety of secondary students' feedback literacy skills is needed. The criteria for the levels of the students' feedback literacy were found functional in the coding of their skills, and these criteria can be used and elaborated on in other studies.

The present study shows that secondary students can be led to develop their feedback literacy skills. The development of their feedback literacy will be of value in their university studies (O'Donovan, 2017) and workplaces (Carless & Boud, 2018). Given that most people will not attend higher education but that feedback literacy skills are relevant to everyone, feedback literacy should be developed in secondary education.

ORCID

Laura Ketonen  <http://orcid.org/0000-0002-2821-0179>

Pasi Nieminen  <http://orcid.org/0000-0001-6019-1054>

Markus Hähkiöniemi  <http://orcid.org/0000-0002-6337-5366>

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.
- Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., Tamm, M., Kamppi, P., Rumpu, N., Hietala, R., & Immonen, J. (2019). *Että tietää missä on menossa. Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa*. The publications of the Finnish Education Evaluation Centre.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice* 25(6), 551–575.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. [Database] <https://doi.org/10.1191/1478088706qp0630a>
- Boud, D., & Molloy, E. (2013). What is the problem with feedback? In D. Boud and W. Molloy (Eds.), *Feedback in Higher and Professional Education* (pp. 1–10). Routledge.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, 35(5), 551–564.
- Crane, L., & Winterbottom, M. (2008). Plants and photosynthesis: Peer assessment to help students learn. *Journal of Biological Education*, 42(4), 150–156. <https://doi.org/10.1080/00219266.2008.9656133>
- Davis, N., Kumtepe, E. G., & Aydeniz, M. (2007). Fostering continuous improvement and learning through peer assessment: Part of an integral model of assessment. *Educational Assessment*, 12(2), 113–135. <https://doi.org/10.1080/10627190701232720>

- Delva, D., Sargeant, J., Miller, S., Holland, J., Alexiadis Brown, P., Leblanc, C., Lightfoot, K., & Mann, K. (2013). Encouraging residents to seek feedback. *Medical Teacher*, 35(12), e1625–1631. <https://doi.org/10.3109/0142159X.2013.806791>
- Foley, S. (2013). Student views of peer assessment at the International School of Lausanne. *Journal of Research in International Education*, 12(3), 201–213.
- 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. <https://doi.org/10.1016/j.learninstruc.2009.08.007>
- Gravett, K., Kinchin, I., Winstone, N., Balloo, K., Heron, M., Hosein, A., Lygo-Baker, S., & Medland, E. (2020). The development of academics' feedback literacy: Experiences of learning from critical feedback via scholarly peer review. *Assessment & Evaluation in Higher Education*. Advance Online Publication, 45(5), 651–665.
- Han, Y., & Xu, Y. (2019a). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education*. Advance Online Publication, 1–16. <https://doi.org/10.1080/13562517.2019.1648410>
- Han, Y., & Xu, Y. (2019b). The development of student feedback literacy: The influences of teacher feedback on peer feedback. *Assessment & Evaluation in Higher Education*, 45(5), 680–696.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. [Database] <https://doi.org/10.3102/003465430298487>
- Hey-Cunningham, A. J., Ward, M.-H., & Miller, E. J. (2020). Making the most of feedback for academic writing development in post-graduate research: Pilot of a combined programme for students and supervisors. *Innovations in Education and Teaching*. Advance Online Publication. <https://doi.org/10.1080/14703297.2020.1714472>
- 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.
- Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63–76.
- Ketonen, L., Häkkinen, 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, 1465–1484. <https://doi.org/10.1007/s10763-019-10030-3>
- Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257–275. <https://doi.org/10.1007/s11251-011-9177-2>
- McConlogue, T. (2015). Making judgements: Investigating the process of composing and receiving peer feedback. *Studies in Higher Education*, 40(9), 1495–1506.
- McLean, A. J., Bond, C. H., & Nicholson, H. D. (2015). An anatomy of feedback: A phenomenographic investigation of undergraduate students' conceptions of feedback. *Studies in Higher Education*, 40(5), 921–932. <https://doi.org/10.1080/03075079.2013.855718>
- Minjeong, K. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education*, 34(1), 105–114.
- Molloy, E., Boud, D., & Henderson, M. (2019). Developing a learning-centered framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527–540. <https://doi.org/10.1080/02602938.2019.1667955>
- O'Donovan, B. (2017). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education*, 74(4), 617–633.
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown, & L. R. Harris (Eds.), *Human factors and social conditions of assessment* (pp. 1–39). Routledge.
- Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133–156. <https://doi.org/10.1007/s10212-015-0282-5>
- 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. <https://doi.org/10.1016/j.stueduc.2013.10.005>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International*, 49(1), 31–40.
- Tasker, T., & Herrenkohl, L. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education*, 27(1), 35–59. <https://doi.org/10.1007/s10972-016-9454-7>
- To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*, 44(6), 920–932.
- 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). 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(4), 1161–1174. <https://doi.org/10.1016/j.compedu.2006.01.007>
- 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. *Journal of Computers in Mathematics and Science Teaching*, 31(4), 433–465.
- Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21(4), 506–519. <https://doi.org/10.1016/j.learninstruc.2010.08.002>
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, 20(4), 280–290. <https://doi.org/10.1016/j.learninstruc.2009.08.010>
- 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. <https://doi.org/10.1016/j.learninstruc.2009.08.004>
- William, D. (2012). Feedback: Part of a system. *Educational Leadership*, 70(1), 30–34.
- Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). It'd be useful, but I wouldn't use it: Barriers to university students' feedback seeking and reciprocity. *Studies in Higher Education*, 42(11), 2026–2041. <https://doi.org/10.1080/03075079.2015.1130032>



III

HOW DO LOWER-SECONDARY STUDENTS EXERCISE AGENCY DURING FORMATIVE PEER ASSESSMENT?

by

Laura Ketonen, Pasi Nieminen & Markus Hähkiöniemi

Under review (Educational Assessment).

Request a copy from the author.