Peer assessment in high school mathematics class

Using peer assessment as formative assessment

By

Hector A. Santoni Gomez

in partial fulfilment of the requirements for the degree of

Master of Science
in Science Education

at the Delft University of Technology,

Date: April 2021
Student number: 4345657

Supervisor:	dr. J.G. Spandaw	TU Delft
Thesis committee:	Prof.dr. M.J. de Vries,	TU Delft
	dr.ir. R. van der Toorn,	TU Delft

TU Delft

Abstract

Peer assessment (PA) is a process that involves students assessing each other's work and providing their peers with (elaborated) feedback. This thesis investigates how an elaborated form of formative PA can be implemented in the mathematics class to promote students' achievement in mathematics. A mixed methods investigation at a third-grade mathematics class was conducted. For this investigation, a teacher's guide and a student guideline were developed. The teacher's guide helps the mathematics teacher in implementing PA in the class. The students' guideline was developed to guide students doing online PA during distance learning due to the COVID-19 pandemic. Using this student guideline can require more effort for the teacher to monitor students' participation and more effort for students to do elaborated online PA. Students were trained with the technique "two stars and a wish" for providing peer feedback. With this feedback technique, students could provide their peers with two reinforcing feedback and one suggestive feedback. The feedback has to first be written in their peer's work and later be elaborated when the work is being returned. As a result of this investigation, implementing PA promoted students to interact and learn from one another during online distance learning. Evidence was found that by implementing PA, students were more involved in their learning process, and students' achievement increased. PA can be used as an effective learning method for students. However, it is preferably for PA to be implemented during physical class, especially when it is implemented the first time.

Keywords: Peer assessment, high school mathematics, peer feedback, elaborated feedback,

Acknowledgement

I would like to express my gratitude to my supervisor Jeroen Spandaw for the constructive comments, remarks, and engagement through the learning process of this master thesis. Furthermore, I would like to thank Professor Marc de Vries and Ramses van der Toorn for giving their time to participate as assessors of my thesis. I also like to thank Walter Wassenaar, the principal of the school, Obed Vreds, the math teacher, and all the students of the third-grade mathematics class at GSR in Rijswijk, for allowing me to do my investigation and gather the necessary data for this thesis.

The year 2020 was difficult for many people as the outbreak of the COVID-19 pandemic hit the whole world. At the beginning of this year, my mom passed away all of a sudden from a heart attack. 2020 is, therefore, an unforgettable year.

I want to use this opportunity to give thanks to my family and friends for all the support I have received. A special thank goes to my wife, Anna Bin Yang, who have supported me all these years, and for our newborn son, Sem, which has brought us a lot of happiness. Finally, I thank God for His wisdom and guidance to help me go through some valleys of life. That said, still, my wife and I are looking forward to moving on to the next stage of our life.

Index

Definitions

Assessment: refers to all those activities undertaken by teachers and by the students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged (Black and Wiliam, 1998, p. 2).

Cooperative learning: "*cooperative learning takes place when students work "cooperatively to accomplish shared learning goals" (Johnson et al.,1998, p. 28)*". (Wiliam and Leahy, 2015, p. 140).

Deep learning: "*is a key strategy by which students extract meaning and understanding from course materials and experiences*" (Warburton, 2003, p. 44).

Elaborated feedback: "*any method that goes beyond providing just the correct response is generally considered elaborated feedback (Kulhavy & Stock, 1989). Elaborated feedback includes, in addition to the correct answer, supplementary information designed to foster deep learning of the target information*" (Finn et al., 2018, p. 104).

Feedback: "*feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way*" Ramaprasad, 1983, p. 4; Taras, 2005, p. 470). For example, information on student's homework assignment used to improve for a future homework assignment is feedback. The system parameter is the student's homework assignment progress.

Formative assessment: "*to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited*" (Black and Wiliam, 2009, p. 9). In other words, assessment during the learning process, in which feedback and points for improvement are central.

Formative feedback: defined "*as information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning*" (Shute, 2008, p. 154).

Summative assessment: the assessment at the end of a process, which determines the development up to that point (Taras, 2005, p. 468).

Peer assessment (updated definition): "*an arrangement for learners to consider and specify the level, value, or quality of product or performance of other equal-status learners, then learn further by giving elaborated feedback and discussing their judgments with peers to achieve a negotiated agreed outcome*" (Topping, 2017, p. 2).

Chapter 1: Introduction

As we all know, mathematics classes usually go like this: the teacher gives instruction in which he or she explains the theory of a certain subject to the students. This explanation can be done on a blackboard or by using a presentation. After the theory has been explained, follow practical exercises. The teacher first explains some exercises in detail, in which each step is made explicit. After this, the students themselves go to work on several similar exercises. After letting the students do these exercises by themselves, the teacher explains these exercises in class. The teacher writes down the steps to indicate to students how to arrive at the correct solution. The theory is further practiced by making more similar (complex) exercises that can be used as homework. The teacher coaches and guides the students in the direction of getting the correct answer by letting them do exercises independently and/or in small groups.

By this traditional way of giving classes in the mathematics class (or in any class), there is a risk that the students will develop a rather passive learning attitude (Rukavina et al., 2012). Is there a way to engage students to be more actively involved in their learning?

A cooperative learning strategy such as *Peer Assessment* (PA), which can be used as a formative assessment strategy (this will be discussed later on in chapter 2), can be used as an active learning method (Amo & Jareño, 2011; Ng et al., 2020; Li et al., 2010). Students are actively involved in the process of assessment (Harris and Brown, 2013) and using assessment information to improve learning (Stiggings, 2001). PA can help learners provide their peers with peer feedback which can help learners to plan their learning, identify their strengths and weaknesses, and target areas for remedial action if PA is effectively implemented (Topping, 2017). Peer feedback may be more effective than teachers' feedback as students are more likely to act on their peers' comments than comments from the teacher (William and Leahy, 2015).

1.1 Goal of the thesis

PA has potential benefits in students' learning, such as increasing students' achievement and providing students with additional sources of feedback (Harris and Brown, 2014; Topping, 2009, 2018; Gurbanov, 2016; Panadero and Alqassab, 2019). Students' achievement is strongly associated with the stimulation of meaningful learning (Schneider & Preckel, 2017). According to Topping (2017), PA can be categorized into two forms: a *simple* and an *elaborated* form, which will be clarified later on in section 2.2. Some authors, like Topping (2018) and Wiliam and Leahy (2015), prefer the implementation of an elaborated form of PA rather than the simple form because the simple form would not provide the student with information about how to improve their work. Therefore, it is interesting to investigate how to implement an elaborated form of PA in the math class.

When searching for research literature about "peer assessment", the works of Keith Topping (1997, 2009) come on top of the list at the search engine Google Scholar, and two of his papers have been cited in total over 3000 times since these papers were written. This

has motivated me to look for the works of Topping about PA implementation. In this thesis, the literature of Topping and others will be used to answer the research questions.

Limited literature research about PA implementation in the mathematics class in high school can be found. When I search in Eric.gov with the term "peer assessment" AND "mathematics" and filtered for high school or secondary school, the results were less than 20 articles. The search results were also similar by using the same search terms in Google Scholar. When I search for “peer assessment” AND “mathematics”, most of the results were research papers in higher education. This statement is also made by Topping (2018) and Harris and Brown (2014).

According to Topping (2018) and Panadero (2016), PA can be used in different curriculum areas or subjects. These authors provide a step-by-step guideline on how to implement PA in the classroom, and they provide examples on how to implement PA in the language class or in the science class when assessing portfolios. An example of classroom implementation of PA in a high school mathematics class, on the other hand, has not yet be found in the literature research. Can these step-by-step guidelines from Topping and Pandero also be implemented in the high school mathematics class?

This thesis will focus on investigating how elaborated forms of PA can be implemented in the mathematics class and what evidence, if any, can indicate students’ learning from peer feedback. A guide will be offered to the math teacher on how to implement a formative form of PA in the mathematics class to promote the practice of peer feedback.

1.2 Research question

Investigating the implementation of an elaborated form of PA in the mathematics class in high school and its effects on students’ achievement, as mentioned in section 1.1, leads to the following research question:

How can elaborated forms of formative PA be implemented in the mathematics class to promote students’ achievement in mathematics?

Implementing PA implies that students will provide each other with peer feedback. Depending on the technique used for students to provide peer feedback, the assessed work could contain some positive and some negative feedback. Topping (2018) promoted in his book the framework of Chi (1996) to identify feedback between four types: reinforcing, corrective, suggestive, and didactic feedback. The four types of feedback will be discussed later in section 2.3. I find it interesting to investigate if these four types of feedback could be found in the feedback provided by students. This leads to the first sub-question:

- What type(s) of peer feedback do students give to their peers by using a specific technique for providing feedback?

After implementing PA in the math classroom, I want to know the experience of the students and the math teacher. This information can be used to improve future implementation in the math class. So, sub-question two is:

- In the views of students and the math teacher, what is the effect of peer assessment on students' learning in mathematics?

As mentioned in the introduction, according to some authors, such as Harris and Brown (2014) and Gurbanov (2016), PA has potential benefits in students learning, such as an increase in students' achievement. After PA implementation in the math class, what are the effect of PA on students' achievement? Can any evidence be found indicating gains in students' achievement? Therefore, the third sub-question is:

- What evidence can be found that indicates gains in students' achievement as an effect of PA?

1.3 Relevance

The Assessment for Learning movement (Assessment Reform Group, 2002; Berry, 2011; Black & Wiliam, 1998) has been promoting formative practices like PA to educators and policymakers to actively involve students in the process of assessment (Harris and Brown, 2013, 2015). This thesis uses the latest developments in PA implementation and also the theories that are relevant for the math class. This thesis can as well contribute to promoting the practice of formative strategies implementation throughout the curriculum for the math class.

According to Topping (2009), teachers need to be trained using PA before implementing it in the classroom. The plan in this thesis is to develop a simple guide with tools and techniques that is possible to train (future) math teachers with PA. The findings from the investigation may also help (future) math teachers to get familiar with this formative strategy.

Chapter 2: Literature review

This research is about applying Peer Assessment (PA) in mathematics class as a formative strategy. Peer Assessment is a form of assessment in which the student is assessed by their peer.

2.1 What is formative and summative assessment?

Teachers commonly use two tools to assess student learning of new knowledge, namely, formative and summative assessment (Dixson and Worrell, 2016). The precise definitions of formative and summative assessment are slightly confusing. In this thesis, the definition of Black and Wiliam (2009, p. 9) for formative assessment, as already mentioned in the descriptions, is: "*to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited*". In other words, assessment during the learning process, in which feedback and points for improvement are central.

Taras (2005) defines summative assessment as the assessment at the end of a process, which determines the development up to that point. In most cases, in summative assessment, the assessment is provided by scores and grades (Taras, 2005). Teachers can use summative assessment "*at the end of a chapter, quarter, or semester to assess and evaluate how much learning students have gained and retained*" (Dixson and Worrell, 2016, p. 157). Summative assessments can contain grades only, comments on how to improve their work only, or grades and comments as feedback (Butler, 1988).

Depending on what the motive of assessing is of the teacher, formative or summative assessment will then be implemented. This motive is important when deciding to apply for a PA form. In the next section, this will be discussed.

2.2 What is Peer Assessment?

As mentioned in the definitions above, peer assessment is "*an arrangement for learners to consider and specify the level, value, or quality of product or performance of other equal-status learners, then learn further by giving elaborated feedback and discussing their judgments with peers to achieve a negotiated agreed outcome*" (Topping, 2017, p. 2). This definition is the new definition for PA that Topping has used since 2017.

As mentioned in the introduction, PA can be categorized, according to Topping (2017), into two forms: a *simple* and an *elaborated* form. The simple form of peer assessment is mainly a type of summative assessment, where teachers have students grade the work of their peers. Elaborated PA is mainly a type of formative assessment, where the assessor (the student assessing) provides the assessee (the student being assessed) with feedback and later engages both in discussions of the assessment (Topping, 2017). During PA, students will act both as assessor and assessee: as assessor, students review their peers' work and

provide constructive feedback, and as assessee, students receive feedback from their peers and may make improvements accordingly to their own work (Li et al., 2010).

Before this new definition, Topping (2003, 2009) used a slightly confusing definition, which could be interpreted as a summative assessment strategy where grading by students is involved. PA was 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*" (O'Donnell & Topping, 1998). From this old definition, some authors, like Wiliam and Leahy (2015) and Panadero et al. (2016), considered peer assessment to be focused on summative assessment. Under the old definition, students were used as substitute teachers where the validity and reliability of students' assessment might be questioned (Panadero et al., 2016).

There has been researched done (Gurbanov, 2016; Panadero & Alqassab, 2019; Li et al., 2020) where PA has been used to compare student's grading each other to the grade teachers would award (Wiliam and Leahy, 2015). These researches about PA implementation have been criticized by Wiliam and Leahy (2015) to be more an academic exercise rather than for PA to be used to advance students' own learning and that of their peers. By this, they are promoting the use of PA as a formative strategy. Wiliam and Leahy (2015) suggest implementing PA not as summative assessment because assigning grades can destroy the positive effect of substantive feedback. Perhaps these arguments could be one of the reasons why, since 2017, Topping started promoting the implementation of an elaborated form of peer assessment rather than the simple form.

2.3 Peer feedback

Peer feedback is a term related to peer assessment, but it does not mean the same thing (Topping, 2017). The learning element of PA is represented by peer feedback (Lui and Carless, 2006). Peer feedback is an essential part of peer assessment, "*it refers to that part of peer assessment which involves peer assessors giving information to their assesses about how they found the strengths and weaknesses of the item that was assessed*" (Topping, 2019, p. 13).

Providing and receiving peer feedback is beneficial to students' learning (Ion et al., 2019; Li et al., 2010). In higher education, Ion et al. (2019) suggested that students benefit more in their learning by providing peer feedback than receiving peer feedback. According to Wu & Schunn (2020), both providing and receiving peer feedback are beneficial to students' learning in secondary school.

The purpose of peer feedback is that students can learn how to improve their work/assignments by following their peer's comments. Using the framework of Chi (1996), it is possible to identify feedback between four types: *reinforcing*, *corrective*, *suggestive*, and *didactic feedback*. In Tseng and Tsai (2007, p. 7), these types of feedback are defined as follows:

- **Reinforcing Feedback:** This type of feedback is given when the provided solution is correct. Recognition or positive feelings are then expressed by using, for example, positive or supportive expression.

- **Didactic feedback:** This is feedback given by the assessor that can be a bit lengthy explanation with a "lecturing-like" tone to help the person being assessed on a solution.
- **Corrective feedback:** This is given to indicate an incorrect starting design or misuse of information used by the student. Therefore, if the incipient information or assumptions are misinterpreted and then used to process the statement.
- **Suggestive feedback:** The assessor provides advice on how to solve a problem. It can be a hint or a warning that there is a problem without telling exactly what the problem is or how the problem can be solved.

When participating in elaborated forms of PA, students will provide each other with peer feedback. This feedback can be positive or negative. Negative feedback indicates where the student's work needs improving (Topping, 2018). Students are not skilled in providing their peers with feedback (Wiliam and Leahy, 2015). Wiliam and Leahy (2015) argue that when students need to provide each other with peer feedback for the first time, it is advisable to use the technique "two stars and a wish". This technique helps students to provide more positive feedback than negative feedback. The "two stars" meaning two positive (reinforcing) feedback, and a "wish" meaning a suggestion to improve the work. This technique provides a structure to students for offering their peers sensitive and constructive feedback (Wiliam and Leahy, 2015).

Another feedback technique that can be used when students provide their peers with feedback is the "ABC feedback technique" (Wiliam and Leahy, 2015). This technique helps students to respond to their peer's work by: (A) Agreeing with their peer by highlighting the areas of agreement; (B) Building up their peer's work by providing suggestions for how the work could be strengthened; (C) Challenging their peer by providing suggestions for additions or more important improvements in their peer's work. With this technique, students can provide more critical/negative feedback than positive feedback.

2.4 Why implementing PA in the mathematics classroom?

According to Wiliam and Leahy (2015), cooperative learning is a specific aspect of formative assessment. By the definition of Topping (2017) about elaborated PA, cooperative learning is a form of PA as a formative type, as mentioned above. When students learn in a cooperative form, scores of students' achievement can be 0.66 standard deviations higher than those students learning competitively and 0.63 standard deviations higher compared to students learning individually (Johnson et al., 1998). Therefore, PA as a formative form can significantly improve students' achievement (Topping, 2003).

Noonan & Duncan (2005) mention four reasons for implementing PA in class, namely: (1) to increase student involvement in the learning process, for example, by taking on teaching tasks, (2) to promote social interactions and confidence in others, (3) to facilitate individual feedback and (4) to help students focus on the process and not only on the product (i.e., learning to learn and not just grading). These are interesting claims. However, there is no evidence on how Noonan & Duncan (2005) came up with these claims in their research. One can only wonder if these reasons are more than based on the authors' intuition.

Topping (2003) says that the nature and purpose of assessments affect many factors of student learning performance, including anxiety and perceived controllability. Depending on the social, cultural, and educational level, assessments may be accompanied by hesitation and uncertainty. PA has been shown to help reduce this anxiety and make assessments stress-free to a degree (Gurbanov, 2016). Giving positive feedback first can reduce the assessed student's anxiety and improve the acceptance of negative feedback (Topping, 2003).

Peer assessment can be implemented in the mathematics class where products to be assessed include writing, oral presentations, test performance, portfolios, or other skilled behaviors (Topping, 2018). These assessments do not have to be summative nor graded by students (Topping, 2018; Panadero et al., 2016).

Implementing PA as a formative strategy can help students understand the assessment process and involve students in assessing and reflecting on their work and their peers effectively (Taras, 2009) if PA is properly organized. Topping (2003, 2009, 2019) repeatedly claims that students with less skill at assessment but with more time to do the asssessment can produce an equally reliable and valid assessment as the teacher. Peer feedback can be obtained in greater volume and is quicker than the teacher's feedback, which can compensate for any quality disadvantage (Topping, 2003).

The next section will discuss the steps to take for organizing PA effectively according to Topping (2017, 2018) and Panadero et al. (2016).

2.5 Organizing PA

Peer Assessment appears to be intuitively easy to use in class. Perhaps one would think that students can review the work of their peers without training or guidance. However, this appears not to be the case. There are some considerations to give attention to before and while implementing PA.

Before implementing PA, it is important that there is a good relationship between the teacher and the students participating in PA-activity, and between a student and his classmates (Harris & Brown, 2013). A good atmosphere in the classroom creates space for students to dare to make mistakes and see this as an opportunity to learn (Wiliam and Leahy, 2015). Besides, this encourages students to reflect honestly on their learning process.

At the beginning of implementing PA, there will be no time savings in the short to medium term as the implementation of PA takes time for organization and training (Falchikov, 2001); Topping, 2003). Implementing PA in the classroom without training or guidelines for the teacher makes the assessment less reliable (Lawrence, 1996; Pond et al., 1995; Topping, 2003). The timescale of the PA activities must be clearly communicated in advance to students (Topping, 2009 and 2003). Students (and especially young children) require clear structure and support during PA (Meusen and Joosten, 2010).

For the organization of peer assessment in the classroom, the important planning concerns evident in the literature (Webb & Farivar, 1994; Panadero et al., 2016; Topping, 2003, 2009,

2017) will be used. This will provide a step-by-step guide for implementing PA for the first time. According to Topping (2017, 2018) and Panadero et al. (2016), the steps to be taken are as follow:

1. Collaborate with colleagues
2. Clarify the purpose of PA with the participants
3. Clarifying the assessment criteria
4. Divide participants
5. Provide guidelines, checklist, and/or tangible scaffolding
6. Specify activities and timetable
7. Provide training on Peer Feedback (Provide training, examples, and practical exercises)
8. Monitor and coach
9. Examine the quality of peer feedback
10. Evaluate and provide feedback

2.6 Schools during COVID-19

In December 2019, a contagious virus emerged in China, which has the name COVID-19, most commonly referred to as "coronavirus" (RIVM publication, accessed 6 July 2020). This virus broke out to the whole world and became a global pandemic (COVID-19 pandemic, Wikipedia, accessed 6 July 2020). On 27 February 2020, this virus reached the Netherlands with devastating consequences to the public sector, which led to an "intelligent lockdown" and social distancing of 1,5 meters (COVID-19 pandemic in the Netherlands, Wikipedia, accessed 6 July 2020). From 15 March 2020 to 2 June 2020, high schools remained mostly closed for students, and schools needed to improvise their teaching methods to distance learning, if possible, through an online platform (VO-raad, accessed 6 July 2020).

On the website of the Association of schools in secondary education (VO-raad, accessed 6 July 2020), the latest news articles about coronavirus and education can be found. After 2 June 2020 until the summer vacation, high school students were again allowed to attend classes at school. However, there were some strict social distancing and hygiene measures to keep (Lesopafstand, accessed 6 July 2020). Not all students could at the same time attend school because of the strict social distancing measures. Many schools decided to organize a combination of online distance learning and learning at school. Hopefully, after the summer vacation, students could normally attend school without keeping social distance.

2.7 Online peer assessment

Online peer assessment can, in this situation, be applied to promote learning. According to Lu and Law (2012), online peer assessment has several advantages over face-to-face assessment, which can be in the classroom.

An example of these advantages is anonymity in marking and giving feedback. Besides, using a computer for PA makes it easier for teachers to monitor the participation and progress of students when using online assessment systems (Topping, 2018). Examples of these systems are "NetPeas" and "Group Support System" (Lu and Law, 2012). Another great advantage of

online PA is the flexibility which allows students to assess their peer's work outside of fixed class schedules (Topping, 2018).

A disadvantage for online PA is that the student that assesses cannot see how pleased or upset their peer might be in response to the assessment (Topping, 2018). A camera and microphone can facilitate getting more in touch with the peer while using videoconference calls.

Chapter 3: Methodology

For this thesis, mixed methods research was conducted where the focus is more on quantitative analysis than on qualitative analysis. The empirical test took place in a third-grade mathematics class located at middle school GSR in the town Rijswijk, The Netherlands. The mathematics class took place several times at school but most of the time online, because of the Covid-19 school measures in The Netherlands.

The mathematics teacher agreed to test peer assessment while he discusses the chapter about quadratic equations. Students practiced with peer assessment by correcting each other's homework and providing elaborated peer feedback.

Participants:
The mathematics teacher has nine years of teaching experience and has been working at this school for approximately two years. For this research, I choose to work with this teacher because one year ago, we used to be colleagues, and he offered to help me with my research thesis.

The class contains a total of 21 students. The sample for the PA activities consisted of 10 students, and the remaining 11 students did not participate in the PA activities. At this school, GSR in Rijswijk, it was not possible to find a class level with a parallel class to use for this investigation. This is why this class had to be divided into two groups. The number of students in these two groups is very small, so the data obtained from an experiment would have no statistical significance. On the other hand, this research is conducting an empirical test to see if there is any effect on students' achievement by implementing PA. By getting an idea of the effect size, we can then estimate how large the sample should be to demonstrate significance.

The students participating in PA were selected by their math teacher. These students became the PA-group. The remaining 11 students not participating in the PA activities will be called the "control group". The teacher also matched students in a fixed duo for them to work together during all PA activities. According to Topping (2017), this pair matching is possible for PA and preferably if the teacher knows the students well and can match pairs based on ability.

Activities at school with the teacher and the PA-group:
At first, the teacher introduced the concept of PA. For this research, a manual in the form of guidelines was prepared beforehand to guide teachers in implementing PA in the mathematics class. This guide was given to the teacher for preparation. See the separate document "Docentenhandleiding voor Peer Assessment implementatie in de wiskundeles" (in Dutch) for the teacher's manual.

A schedule was also made in order to plan the PA activities:

1) **Pilot test**. The student guidelines were tested with two students during a pilot test. From this test, corrections and suggestions to these guidelines were collected. These guidelines were attached to each homework assignment.

2) **Introduction of PA to the students**. Students were introduced to the concept of PA, and also, they were informed about the upcoming activities of PA during the math class. The concept of feedback giving was also introduced to students by a PowerPoint presentation I prepared which was presented by the teacher (see the separate document "*Presentation Feedback*" in Dutch).
3) **Practicing with PA**. Students practiced giving peer feedback by doing (online) PA on each other's homework. During the online classroom, students first needed to send their homework to the teacher and their group member. The teacher then sends each student the solutions for them to use during PA. This practice took place twice at school and four times during online classrooms. So, in total, students practiced six times with PA. See Table 1 Data collection schedule.
4) **Test**. The chapter that the teacher taught ended with a test in which the grade did not count. The PA-group and the control group both made this test and were later assessed by the teacher. The test results are presented in Chapter 6.

Student guideline for PA:
The prepared student guideline guides the student step-by-step to perform the homework assignment and do elaborated (online) PA after doing homework. The template of the student guideline can be found in Appendix 1. The steps of this guideline are here explained:

- Steps 1 to 3 are the steps students need to undertake on their own homework assignment (this is indicated in blue color). First, the student needs to do their homework before the next class and send it to the teacher and group member.
- Steps 4 to 6 are steps students need to undertake on their group member's homework assignment (this is indicated in red color). Students review their peer's homework during the next class and write feedback according to the technique "two stars and a wish" promoted by Wiliam and Leahy (2015).
- Step 7 is the crucial step in this assignment because, at this step, students were involved in providing each other with verbal explanations and discussing findings. This step is what Topping (2018) considers as elaborated peer feedback and is the crucial step during PA.
- Step 8 promotes learning from feedback by asking students to write down their self-reflection on the peer feedback and reviewed homework. Students can use their self-reflection to improve the next homework assignment.
- Step 9 is for administration purposes for the teacher. With these assignments, the teacher can monitor this PA activity and evaluate the practice of PA. During the next class, the teacher can give students feedback on improving this PA activity.

Steps that were taken implementing PA in the mathematics class:
These are the steps taken for conducting an elaborated form of PA in a high school mathematics class.

First, a teachers' guide was developed to help a math teacher understand the process of PA implementation and to prepare the teacher to conduct PA implementation in the class. The

teachers' guide provides the teacher step-by-step instructions through PA implementation (see the separate document "*Docentenhandleiding voor Peer Assessment implementatie in de wiskundeles*" (in Dutch) for the teacher's manual). These steps are the steps already mentioned in Chapter 2.5, wherein these steps are discussed in detail.

Second, students were introduced to the concept of PA and feedback. For this introduction, a PowerPoint presentation was made, which the teacher used during the online class for the PA-group. Besides introducing and explaining the concept of PA and feedback, the four types of feedback from the framework of Chi (1996) were also introduced and explained. At the end of the explanations, an exercise about providing feedback was given to the students. The teacher explained this exercise later in the class.

Third, a students' guideline, which was developed for the homework assignments during distance learning, was first tested during a pilot test with two students before the PA activities took place. This pilot test took place during distance learning. These two students followed the instructions, and they performed the PA activity as instructed in the homework assignment. As feedback from this pilot test, there are two points suggested to be considered: 1) the instruction in the assignment were consider too much work to read; 2) contacting each other by telephone or by chat in order to provide elaborated peer feedback was considered too much effort to do. The students' guideline was not adjusted to these two feedbacks (see section 7 for the discussion). After the pilot test, the teacher introduced and explained the steps of the students' guideline for doing an online PA and how to provide elaborated feedback as indicated in the assignment.

Fourth, students practiced with an elaborated form of formative PA during homework assignments as indicated in the students' guideline. Students made their assigned homework first. During the next class, the teacher provided the students with the answer's sheets, which were used to correct the homework. Besides correcting the homework from their group member, students also provided with written feedback. Afterward, the duos discussed and elaborated on their corrected work and written peer feedback and made improvements to their own homework accordingly. This PA activity took place twice at school and four times at home following online distance learning due to the national lockdown.

Fifth, an evaluation of the PA activities was done in the form of a questionnaire and individual interviews. The results of the questionnaire and interviews are presented in chapter 6.

Notice that most of the steps about implementing PA, as described in section 2.5 above, are present in these actions that were taken during the investigation at school. A step that is not completed during the experiment at school, due to the lack of time, is providing feedback back to students (step 10 in section 2.5).

3.1 Data collection

From the experiment at school, quantitative and qualitative data were collected using different methods as follows in the schedule (see Table 1 Data collection schedule). These data were collected in order to answer the research questions:

1) An online **questionnaire** in Dutch with open and closed questions for the PA-group. Students filled out this questionnaire which was focused on their perception during the PA activities and on their roles as assessor and assessee (which was reviewing their group member's work and giving- and receiving feedback as described in section 2.2).

   The questionnaire contains ten questions: 5 open questions where students chose from a selection of responses (quantitative data); 5 closed questions where students could write their answer (qualitative data). The questions for the questionnaire can be found in Appendix 4. The guidelines from Lambert (2012, p. 118-121) for developing this questionnaire were followed.

   The questionnaire will provide data to help answer sub-question 2.

2) Three semi-structured face-to-face individual **interviews** were done in Dutch. Two interviews were with two students separately, and one was with the teacher. As suggested by Lambert (2012, p. 123), the structure of these interviews was: introduction, simple start, meaty middle, and rounding off. The questions during the meaty middle and rounding off are only presented. After the introduction and simple start, the middle part of the interview questions to the students was mainly to ask about their experience during the PA-activities, reviewing their group member's work, and providing- and receiving feedback.

   The interview with the teacher was on the last day of the experiment at school. The interview questions were mainly about his experience with the teachers' manual, which was provided, his perspective of the PA activities, and his opinion of implementing PA in a future class.

   The interview questions can be found in Appendix 4. All interviews were voice-recorded and then transcribed. These interviews provide data to help answer the main question and sub-question 2. Coding is used to analyze this data (see Analysis of data).

3) **Document analysis**: Document analysis was used to analyze 20 documents of the homework assignments. This data will help answer sub-question 1.

4) **Testing**: The teacher assessed 20 tests made by both groups. These tests will provide data to help answer sub-question 3.

The following table is presented when the data was collected.

Table 1: Data collection schedule

Date	Nature of data collection	Sample	Duration	Notes
27/May	Pilot test	Two students	30 min	The pilot test after introduction PA through online teaching
29/May	Observation of mathematics lesson online teaching	Teacher	80 min	Written notes taken
03/Jun	Observation of mathematics lesson Assignment PA	PA-group and teacher PA-group	120 min 40 min	In the class, two class hours Written notes taken Two apart groups of PA
05/Jun	Observation of mathematics lesson online teaching	Teacher	80 min	Written notes taken
10/Jun	Observation of mathematics lesson online teaching	Teacher	80 min	Written notes taken
12/Jun	Assignment PA Observation of mathematics lesson Individual Interview	PA-group PA-group and teacher Two students participating in the PA-group	30 min 50 min 10 min	In a different classroom at school Written notes taken Two apart interviews Voice-recorded
17/Jun	Observation of mathematics lesson online teaching	Teacher	80 min	Written notes taken
19/Jun	Collection of teacher assessment of the test Questionnaire Individual Interview	Teacher PA-group Teacher from observed mathematics lessons	65 min 15 min 22 min	Written notes taken Online questionnaire Interview voice-recorded

3.2 Analysis of the data

Coding:
The transcripts data of the interviews were analyzed by using the process of coding. Lambert's (2012, p. 170) and Bryman's (2016, p. 581-584) guidelines for coding were followed.

First, the transcript data was copied to a new document for analysis. In this document, five columns were made for the *Dialog*, *Speaker*, *Text*, *Analysis*, and *Notes* (translated from Dutch: Dialoog, Spreker, Tekst, Analyse, en Notities). Under these columns are the following contain:

- *Dialog*: number of the dialog of the interview is presented.
- *Speaker*: the role of the speaker is at the moment.
- *Text*: the transcript data of the interview.
- *Analysis*: the labels (words or abbreviations) are presented, which represents a code that relates to the corresponding transcript data. These codes are part of a coding system.
- *Notes*: here, the extra thoughts and ideas of the text are presented.

Lambert (2012, p. 170) calls this process "coding", which is a process of classifying chunks of the interview data into key themes or headlines (Robert-Holmes, 2011, p. 186-187). An example of how coding was done with the interview transcript of student 1 is presented in Figure 1. The same codes were used for coding across all transcripts.

Dialoog	*Spreker*	Tekst	Analyse	Notities
⋮				
3	*Interviewer*	…… Maar wat vond je zelf van feedback geven aan haar?	Feedback geven	
4	*Student 1*	Nou, ik vind feedback meer geven nooit zo heel moeilijk. Ik kan wel snel overzien wat je ongeveer, dat zou iets beter kunnen. Dus, ik vind het prima, niet zo moeilijk.	Feedback geven Capaciteit Feedback geven	Feedback geven vindt student 1 geen probleem
5	*Interviewer*	Okay, goed. En feedback ontvangen?	Bevestiging Feedback ontvangen	
6	*Student 1*	Dat vind ik trouwens heel handig, want dan weet je waar je ongeveer zit. Dan je denk van "ooohhh, maar als ik dat had gedaan" of "dat heb ik het goed gedaan, dus dat moet ik blijven doen", weet je wel? Dat soort dingen.	Acceptatie feedback ontvangen Beseffen leren	Feedback ontvangen vindt student 1 veel belangrijker

Figure 1: Example of coding of a segment interview with student 1

Next, all coded data were reorganized according to the codes by doing the following steps proposed by Lambert (2012, p. 171): 1) Grouping the data; 2) Examine data; 3) Compare themes and perspectives; 4) Digging deep into the data. After analyzing the data, the findings were used to answer the research question.

Document analysis:
Data from the homework assignment was gathered. Students were asked to fill in the assignment, which was provided with the homework. The process of coding was used to analyze the data gathered by the filled-in assignment. This coding process is similar to the coding process to transcripts, as explained above.

First, the provided peer feedback in the assignments was written over to a Word document for analysis. In this new document, a column for coding (column name "analyze type feedback") and a column for notes (column name "notitie") were added. The codes (in Dutch) used for coding the peer feedback were as follow:

- Pos= Positive feedback
- Neg= Negative feedback
- V= Reinforcing Feedback (in Dutch: *Versterkende feedback*)
- D= Didactic feedback (in Dutch: *Didactisch feedback*)
- C= Corrective feedback (in Dutch: *Correctief feedback*)
- S= Suggestive feedback

Next, the data was coded Pos or Neg, and then the data was further coded (as V, D, C, and S) based on the definitions from Tseng and Tsai (2007, p. 7) for each type of feedback as explained in section 2.3. An example of how coding was done with the gathered peer feedback of a student is presented in Figure 2.

Student 9	29 mei Peer feedback	Analyse type feedback	12 juni Peer feedback	Analyse type feedback	Notitie
Ster 1	Je snapt het hoe je moet berekenen	Pos V	Je snapt de stof goed!	Pos V	Leerling blijkt niet geleerd te hebben van vorige feedback. Blijkbaar is zijn/haar huiswerk niet volledig
Ster 2	Hartstikke mooi	Pos V	Geen slordigheidsfout jes	Pos V	
Wens	Nauwkeuriger opschrijven van de stappen	Neg D	Bij een schets hoef je niet alles uitgebreid te schrijven	Neg S	
Zelfreflectie	Echt alles maken, nauwkeuriger zijn		Dat ik het moet maken		

Figure 2: Example of coding peer feedback

Finally, the codes were sorted, and the feedback was sorted based on its type, as presented in Appendix 2. These are the results for sub-question 1, which will be explained in section 6.1.

Testing:
The data from the test scores obtained was evaluated by comparing the difference in the average scores of both groups and see if there is any significant effect. For this evaluation, the Effect Size of the average scores will be calculated by using the effect size index Cohen's d (Cohen, 1988). The Cohen's d will provide a measure of the size of the effect PA has on student achievement.

The value of Cohen's d can be calculated by the following simple equation from Rosenthal and Rosnow (2008, p.385), which uses the statistical t -value from a two-sample size t-test. The weighted Cohen's d is presented in (1).

$$\text{Cohen's } d = \frac{t(n_1+n_2)}{\sqrt{(n_1+n_2-2)(n_1 \cdot n_2)}} \qquad (1)$$

where t is the t-value and n_i is the sample size of group i.

Cohen's conventions for effect size (Cohen, 1988):

- Small effect for $d = 0.2$
- Medium effect for $d = 0.5$
- Large effect for $d = 0.8$

To test the statistical significance of the Cohen's d, the 95% Confidence Interval of the Cohen's d will be estimated by the formula of Hedges and Olkin (2014):

95% CI for Cohen's d: $[\mathrm{d} - 1.96 \times \sigma(\mathrm{d}),\ \mathrm{d} + 1.96 \times \sigma(\mathrm{d})]$, (2)

where $\sigma(d) = \sqrt{\frac{n_1+n_2}{n_1 \times n_2} + \frac{d^2}{2(n_1+n_2)}}$ (3)

is the Standardized Error (SE).

Chapter 4: Validity and reliability

The literature was tested in an actual high school. To strengthen the validity and reliability of my investigation, I took suggestions from the math teacher for correcting the teacher's guide to be more reader-friendly and the steps to implement PA to be more precise. A student guideline was developed for online peer assessment.

To lessen the threat of unclarity in the student guideline, I took suggestions from the math teacher to adapt this guideline to be more reader-friendly for students. I also added color text to indicate action steps and a picture logo to indicate pasting the picture of the homework. For the use of this student guideline, a pilot test was conducted with two students before using to train students with PA. Triangulation has been used to answer sub-question 2, where combinations of the four methods were used to gather data at school.

The articles for the literature study were searched through reliable search engines, such as Google Scholar and ERIC. The keywords used were "peer assessment", "peer assessment" AND "in high school", "Peer assessment" AND "assessment for learning", "summative peer assessments", "peer assessment training", "online peer assessment" AND "high school". The search was filtered for articles published between 2005 to 2019. The articles were selected that are relevant to this research and the topic and preferably peer-reviewed articles.

Chapter 5: Ethics

Before I started with this research at school, I expressly asked the school director for permission (see Figure 4 in Appendix 3). From the math teacher, where I did my investigation, I received the news that the director permitted me to start with my investigation at the school (see Figure 5 in Appendix 3).

When I arrived at the school, I was introduced by the math teacher to his colleagues, where I took the opportunity to inform them about my project at their school. For my investigation, I obtained consent passively from the participant students in the PA-group. The math teacher nor I received any objections from the participant students.

When collecting data during my research, I made sure that the students' identities in the class remain anonymous. During the recording of the interviews, the names of the interviewees were not mentioned, and when transcribing the audio-recording, only the roles of the interviewees were written.

Chapter 6: Results and Analysis

The research questions are answered in this chapter by using the gathered data from the investigation at the school. First, the three research sub-questions are answered, and then the main research question is answered. In each section of the sub-questions, the results are presented and followed by the analysis.

6.1 Sub-question 1: What type(s) of peer feedback do students give to their peers by using a specific technique for providing feedback?

Data obtained from 20 written homework assignments were analyzed to answer this question. As mentioned in the literature research in Chapter 2, feedback can be identified in four types according to the framework of Chi (1996). After the analysis, the feedback provided by students in Dutch to their peers was sorted into these four types of feedback, namely: *reinforcing*, *corrective*, *suggestive*, and *didactic feedback.* The sorted students' feedback can be found in Appendix 2.

In the assessed homework, students provided their peers with written peer feedback by using the technique "two stars and a wish" promoted by Wiliam and Leahy (2015). All the "stars" given were positive reinforcing feedback, in total 31 times. Only four times students did not provide any written feedback in the assignments. For the "wish", 15 times students provided suggestive feedback. Corrective feedback was provided two times, and Didactic feedback was four times.

Analysis:

As the results have shown, students did not show having difficulty providing their peers with positive reinforcing feedback. For improving feedback (negative feedback), students tend to provide more suggestive feedback to their peers than didactic and corrective feedback. Students use more a constructive tone in their feedback with no use of demeaning words or tone. These results from this mixed methods experiment show that the technique "two stars and a wish" can be used for providing a structure to students that support building students' skills to give their peers sensitive and constructive feedback. This technique has helped the PA-group to provide peer feedback, especially this time when PA is implemented for the first time.

6.2 Sub-question 2: In the views of students and the math teacher, what is the effect of peer assessment on students' learning in mathematics?

Data obtained from the questionnaires and interviews are used to answer this question. The transcripts of the interviews can be found in the separate document "*Interviews op school*" and the results from the questionnaire in "*Resultaten vragenlijst PA groep*", which are in Dutch.

In the views of students:

The majority of students who participated in the PA activity were neutral (50%) or enjoyed (10%) the PA activity compared with the participants that did not enjoy it (40%). The reason for this result is that not all participants finished their homework earlier before attending class. Some of the students that did not do their homework regularly were struggling with their motivation to do homework as Student 1 said during the interview: *"I notice that my motivation fluctuates a lot. So, sometimes I suddenly have that I can do something and sometimes I cannot, and I have days when I do nothing at all. Just no motivation"*.

Students were asked if reviewing their group member's homework contributed to their learning mathematics. The results are mixed. Some students did not learn much from reviewing, and reviewing their peer's homework took a lot of effort (Interview Student 1*:" It's okay to do, but yeah, I don't feel like I learned much out of reviewing. Sometimes I thought, "I could have done it also that way". But on the other hand, it was more work for me to review").* On the other hand, others did learn from reviewing their group member's homework (Interview Student 2:" *You learn more from reviewing than doing the homework yourself and, for the rest, do nothing else with it").*

Besides reviewing their group member's homework during the PA activity, students provided each other with formative peer feedback. From the questionnaire, 60% of students indicated that they did not learn to improve their homework from reviewing and giving feedback to their peer's homework. In comparison, 40% indicated to have learned to improve from reviewing and giving feedback. Students have learned the most from these following skills by giving feedback: to give (positive) feedback besides corrections, to review their peer's work critically, how well their peer's homework was done compared to their own. For some students providing peer feedback is not difficult, especially if they are friends (Interview Student 1: "*Well, I have a good friend, so we give quickly each other a bit of support*"). One student chooses to give almost always the same feedback to their group member (Interview Student 2: "*Aaah yes. Well, it's almost always the same. I give feedback that he should better check his answers from my group member*").

Students were also asked if they agree with the feedback they have received from their groupmate. From the questionnaire, 60% agree with the feedback received from their group member, 20% more or less agree, and 20% disagree. Students want to improve the following skills after receiving feedback: to write more precisely, to do the homework better and learn from it, to check the answers before handing in, to write down all calculations. One student indicated that receiving peer feedback can be "very handy" as ways of improvement are being provided by their peer member (Interview Student 1: "*I find that very handy because then you know where you go with it. That you think "ooohhh, but if I had done that…" or "I did well, so I have to keep doing that").* On the other hand, another student claims to know already how to improve his homework before feedback is giving (Interview Student 2: "*I always know a little bit about what I did wrong anyway. This is usually the case because I have a bit of time that I cannot check my answers and that is often the problem*").

In views of the math teacher:

When the teacher was asked about his perspective of student's enjoyment with the PA-activities, his response was as follow:
"Well, I noticed how fanatically they react to each other. That they really took the time to send really good feedback to each other. And I really like that. You don't see much of "I don't feel like doing this" or anything. Maybe in the sense of "well, it doesn't take extra work". But at the end, something beautiful came out on paper, and you see that a kind of enthusiasm among the students."

During the interview with the math teacher, when asked if the PA-activities contributed to students' learning, this is the teacher's response:
"Look, what you mainly see, what you see very beautifully is…. We tested two students first, and then you saw that in the first step that we roll it out to the group, those two students knew exactly what to do. So, that means, you look at something like a learning curve, you know. Look, it is difficult, that it is really hindering that we have in this period physical lessons and digital lessons. That is hindering. Because this way does not consistently show the students how they should do things. Sometimes they have to email things to each other, sometimes they are here at school. And that is inconvenient for the students. It's much easier for them to say "we always do the same thing. You've seen it one time, "oohh, the same thing next time", some may need another practice, and the third time everyone is fine because they know what to do. And at this moment it is difficult, you have less control, you have less leverage because you do not see the students when they are home. They have to do something at home… So yes, and PA is helpful for the group, it is good, but most of all, it is actually something you should implement when you teach physically. When you are together".

Analysis:

According to students' and the math teacher response, students' attitude toward PA activity as a formative strategy is mixed. Lack of students' motivation for doing homework was a factor that influenced the attitude of participating in the PA activities. Distance learning could have been the cause of this demotivation factor while PA is first time implemented. However, the teacher is satisfied with the participation of the students during the PA activity. The teacher recognized a kind of enthusiasm students put in as they provided satisfactory feedback to each other.

Receiving peer feedback contributed better to improvising student's self-reflection on improving their own homework than by giving peer feedback to their group member. Apparently, most of the students tend to agree with the feedback they received. On the other hand, it might have been that students were trying to avoid entering into a discussion about the received feedback.

In the teacher's view, the PA guidelines added to the homework could better influence the student's learning if the circumstance at school were back to normal. As mentioned already, students could not regularly attend classes at the school. Some days students were present at school. Other days they stay at home and followed distance learning. According to the teacher, this situation hindered the effect of PA in student's learning because "*this way does not consistently show the students how they should do things"*. In this situation, the teacher

has less control to see if students are working properly at home. It would be preferable to introduce students to PA during class and not by distance learning.

6.3 Sub-question 3: What evidence can be found that indicates gains in student achievement as an effect of PA?

In Figure 3, the results from the test are presented. This graph contains only the total score points of each student in the PA-group as in the control group. The students participating in this test were nine from the PA-group and eight from the control group. The maximum score is 17 points.

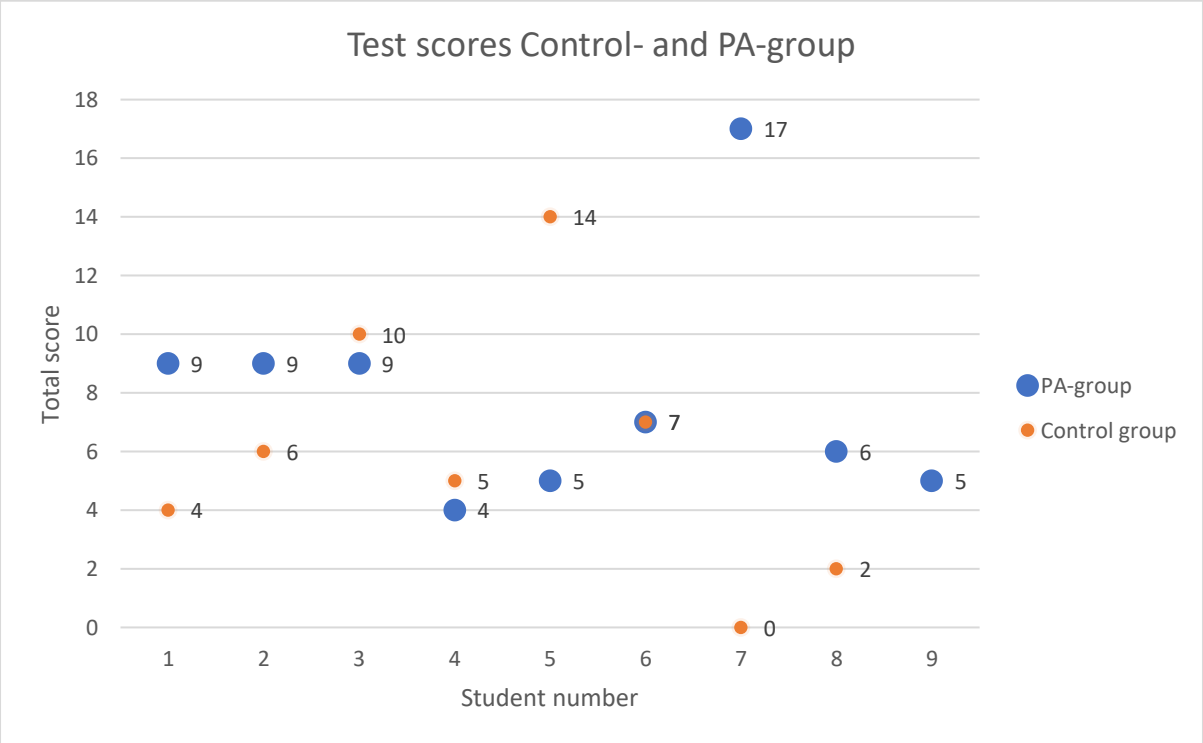

Figure 3: Test results from the Control- and PA-group

Analysis:

From these two groups' test scores (samples) in Figure 3, the mean, standard deviation, and effect size of the means were calculated in MS Excel. The Effect Size is determined by calculating Cohen's d. From testing the statistical significance p-value of the two samples, a two-sample Student t-Test with unequal variances was conducted in MS Excel. See Table 2 for these statistical results.

Table 2: Results from the two-sample Student t-Test

Category	Mean of total score points	Variance
PA-group	7,89	15,3611
Control group	6,00	19,7143
Observations	9	8

t-value= 0,9249 df=14 p-value= 0,3707

Using the result from the t-Test as input for equation (1), the Cohen's d can be calculated, which gives the effect size of d= 0,4784

The sample sizes are $n_1 = 9$ and $n_2 = 8$. Students in the PA-group scored on average (7,89) higher than those in the control group (6,00). To investigate if PA makes any difference in student achievement, the Cohen's d effect size is used. In this case, the effect size (0,4784) is close to 0,5, indicating a medium effect size. In other words, the positive effect of PA on student achievement is medium.

To determine if this positive effect in student achievement is statistically significant, the 95% Confidence Interval is determinant. The 95% Confidence Interval of Cohen's d was calculated by using the result from the effect size as input for equations (2) and (3), which gives as SE for Cohen's d, $\sigma(d)$= 0,4928 and 95% CI [-0,4874; 1,4443]. This interval indicates that Cohen's d can fluctuate between positive and negative values and, therefore, not statistically significant. This means that it is not possible to get with 95% certainty a positive effect size with this sample size in both groups. Therefore, the sample size in both groups needs to be increased. Only in this scenario, PA has shown to have a positive medium effect on student achievement, but replicating this effect is more difficult with these sample sizes.

The t-Test conducted on both samples indicates statistically insignificant because the p-value of 0,3707 is larger than 0,05. This result was already expected because the sample size of both groups is too small for this experiment.

6.4 Research question: How can elaborated forms of PA be implemented in the mathematics class to promote students' achievement in mathematics?

From the results and analysis of the three research sub-questions provided in sections 6.1, 6.2, and 6.3, an answer to the main research question is provided.

Firstly, the teacher has to be trained or guided in PA implementation. The developed teacher's guide for this thesis can be used for PA implementation in the mathematics class. This teacher's guide can guide the teacher through the ten steps for organizing PA, as discussed in section 2.5. According to the math teacher, the teacher's guide/manual is sufficient to guide for PA implementation in the math class. For a busy teacher, a compacted version of this guide is desirable (Interview teacher: "*Yes, I think the teacher's manual was sufficient and pretty clear. But as a busy teacher, I would rather have these as a roadmap rather than a manual*").

When students start for the first time with PA, it is recommended by Wiliam & Leahy (2015) to use the technique “two stars and a wish” for students to provide their peers with two positive feedback (the “stars”) and a suggestion to improve the work for the future (the “wish”). As investigated to answer sub-question 1 in section 6.1, this technique can help students provide a structure that supports building students’ ability to give their peers sensitive and constructive feedback, especially when PA is implemented for the first time.

When PA cannot physically be implemented at school, the teacher can execute this learning strategy during online distance learning. However, the results of the effect of PA in students learning mathematics may vary. The results in section 6.2 show:

- Students experienced demotivation for doing homework during online distance learning and therefore found the experience with PA less enjoyable.
- Receiving peer feedback contributed better than giving peer feedback to their group member.
- The math teacher found it challenging to monitor students during distance learning. In the math teacher's view, PA can be implemented better when students are physically present in class, than through online classes when students are not accustomed to online learning.

Nonetheless, PA can have some positive effect on student achievement than when compared with students not participating in these activities. As a result, found in section 6.3, an effect size in student achievement close to 0,5 can be achieved. However, the result for this particular experiment is not statistically significant because the sample size for both groups is too small.

Chapter 7: Discussion

For the development of the teacher's guide, I mostly used the literature of Topping (2017, 2018), Panadero (2016), and Wiliam and Leahy (2015). I used Topping's literature because his book is entirely about peer assessment implementation and is recently published. Nonetheless, there is more literature to be considered to improve this teacher's guide in other sets than only in the classroom. For example, the literature of Lin (2019) of Theng and Tsai (2007) about online peer assessment implementation could be included as an added section when the teacher's guide is updated. For the current version of the teacher's guide, I explained how PA could be implemented in the classroom, but I developed a student guide that is based on online PA. This student guide I used as an example in the teacher's guide.

Before giving the student's guidelines to the PA-group, a pilot test was first tested with two randomly selected students of the class to try out the guidelines. As feedback from this pilot test, I received that the assignment's text for doing an online PA was too long and could be shortened. However, I did not exclude these steps from the homework assignment because these steps could be used as a reminder of what students were expected to do. At the start of (online) PA, students should read the student's guidelines entirely and follow the instructions. After enough practice, students do not have to read these guidelines anymore.

The results from sub-question 1 in section 6.1 confirm the intended results from the technique "two stars and a wish" as promoted by Wiliam and Leahy (2015). Using this technique, students tend to provide each other with suggestions rather than other types of feedback when negative feedback is given.

As mentioned in section 2.3, in secondary school, both providing and receiving peer feedback are associated with being beneficial to students' learning (Wu & Schunn (2020). On the other hand, the results from section 6.2 indicate that receiving feedback was more beneficial to students than providing feedback. These results do not go by the claim just mentioned. Perhaps this claim by Wu and Schunn (2020) is only appliable to language subjects in high school.

Because of the pandemic, the results from the collected data are affected. A reason for this is that students and teachers are not familiar with online distance learning. According to the literature, PA should promote a sense of ownership, personal responsibility, and motivation in student's learning (Topping, 2018). During this investigation, it was not always that case as Topping suggests. The interviews I collected with the students confirm that students from this class were less motivated to study at home and do homework. The homework was not always made on time. At home, it took the PA-group much effort to contact each other to provide elaborated peer feedback. When students were at school, their participation in the PA activity was feasibly better. The teacher had the feeling of less control over the class during online classes because he had difficulties monitoring students.

During the test, I assisted the teacher in surveilling the PA-group while the teacher was a surveillant for the control group. The policy at school was that no more than ten students were allowed together in one classroom because of the social distance rule during COVID-19. My presence with this group could have affected the test results because the PA-group was more diligent in making the test than compared with the control group. Therefore, conclusions of the effect of PA on students' achievement are not entirely objective.

The effect size 0,4784 of test scores in section 6.3 is not statistically significant. The minimum number of students per group can be calculated using equations 2 and 3 in section 3.2. For this calculation, let's assume that for both groups, the number of students is equal. The left side of the 95% CI equation has to be larger than zero. This leads to a minimum sample size of 35 per group. So, for the effect size of 0,4784 to be statically significant, both the PA-group and control group have to have at least 35 test results. For future research, it is recommendable to take the sample sizes into account in order for the results to be statistically significant.

This research is of small-scale mixed methods research, which was only tested in one mathematic class. On the other hand, this research shows that it is possible to implement PA as a formative strategy in the mathematics class to promote students learning, as the literature claims, which is mentioned in the first two chapters. I hope that the amount of literature about PA implementation in mathematics classes will increase in the near future.

Recommendations from the teacher:

In the interview with the teacher, I asked him if he can give me recommendations for improving this research. His response was as follows:

"*…maybe after a few more times you might be able to take a look with the class of "look, someone submitted this, someone got this feedback". And then reflect the class that process you go through. And then ask the students, for example, "did the feedback you received benefit you", or something. When you hear a student and say "yes, I think that's useful, because ...". Then it becomes more tangible for the class or for those who all participate, how useful this is. Because this is super useful for those who give feedback and for those who receive feedback*." [Translated from Dutch] (See transcript interview teacher line 62).

I agree with his suggestion for a next time implementation. In this case, if I had more time to do this experiment at school and the students were physically present in the classroom, I would for sure consider sharing my reflections on the process of giving feedback by the students. For instance, some reinforcing feedback that students provided could be better specified, "*A lot is right* [Dutch: Veel goed]" (see Appendix 2 first column). This feedback is too vague for the receiver to understand what specific part of his/her homework was done right and what can be improved or should be corrected.

Chapter 8: Conclusions and recommendations

Peer assessment (PA) is a process that involves students assessing each other's work and providing each other with (elaborated) feedback. This thesis investigates how PA can be implemented in math classes to promote students' achievement in mathematics. From the results of the collected data, conclusions and recommendations are here given.

For this thesis, a teacher's guide was developed to guide step by step the math teacher on how to implement peer assessment in the classroom. According to the literature, peer assessment can be implemented in mathematics classes to improve student learning (Topping, 2018). Before implementing peer assessment, students need to be trained on working with PA and practicing giving elaborated peer feedback. A student guideline was developed to guide the student step-by-step to perform elaborated online PA for assessing homework assignments. This guideline was used while students receive online distance learning due to the Covid-19 pandemic.

As a result of this mixed methods research at a third-grade mathematics class in high school, the types of feedback students provided their peers were mostly reinforcing and suggestive feedback. This result was expected from literature for applying the technique "two stars and a wish" promoted by Wiliam and Leahy (2015). This technique can be used for providing a structure to students that support building students' skills to give their peers sensitive and constructive feedback, especially when PA is implemented the first time in the classroom.

PA can be implemented as a learning strategy during online distance learning when physical attendance at school is not possible. However, the intended results may vary when PA is implemented the first time during online distance learning. By this experiment, some students struggled with their motivation to do homework, which impacted the results of the effect of PA. Receiving peer feedback contributed better to improving students' self-reflection on how to improve their homework than by giving peer feedback. According to the math teacher, PA is an effective learning method for students, but it is preferably to be implemented during physical class. At school, the teacher can monitor and guide students more effectively than during online teaching. However, due to the Covid-19 pandemic, teachers and students had to implement PA online with imposed distance learning. The experiment had to adapt to the circumstance.

From the test result of the PA-group and the control group, no statistically significant conclusions of the effect of PA can be made. The positive medium effect of PA on student's performance is not statistically significant for this experiment. The number of test results is too small to be 95% certain of this positive medium effect on student's performance. For this reason, I would recommend further research to be conducted based on a larger number of test results for both groups. To obtain the same positive medium effect in student's performance as from this experiment, the sample size for each group should be larger than 35.

In conclusion, the teacher's guide can be used for guiding the mathematics teacher in implementing PA in the class. The student guidelines for online PA can be used during distance learning. However, the teacher needs to put more effort into monitoring and

evaluating students' participation in providing elaborated feedback to their peers. Due to the Covid-19 pandemic, it was an unfavorable situation to implement PA for the first time in this class, but still, implementing this process promoted students to interact and learn from one another during distance learning. When students start for the first time with PA, it is recommended to use the feedback technique "two stars and a wish" for students to provide their peers with two positive (reinforcing) feedback (the "stars") and a suggestion to improve their work for the future (the "wish"). Even in a small-scale experiment, evidence can be found that by implementing PA, students are more involved in their learning process, and students' achievement can increase. Therefore, this is one method on how an elaborated form of formative PA can be implemented in the mathematics class to promote students' achievement in mathematics.

Recommendations:

Two recommendations for future research are given:

1) To implement PA with a larger population to support the validity of the effect of PA on students' achievement in the math class.
2) To investigate what type of peer feedback students would provide their peers when other techniques for providing feedback are used, such as "ABC feedback".

References

Amo, E., & Jareño, F. (2011). Self, peer and teacher assessment as active learning methods. *Research Journal of International Studies*, *18*, 41-47.

Assessment Reform Group, (2002). Assessment for learning: 10 principles. Retrieved November 1, 2020, from https://www.researchgate.net/publication/271849158_Assessment_for_Learning_10_Principles_Research-based_principles_to_guide_classroom_practice_Assessment_for_Learning

Berry, R. (2011). Assessment reforms around the world. In *Assessment reform in education* (pp. 89-102). Springer, Dordrecht.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*, 139-149.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5.

Bryman, A. (2016). *Social research methods*. Oxford university press.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British journal of educational psychology*, *58*(1), 1-14.

Chi, M. T. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, *10*(7), 33-49.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*. Hillsdale: Lawrence Erlbaum.

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, *55*(2), 153-159.

Falchikov, N. (2001). Learning together: Peer tutoring in higher education. *Psychology Press*.

Finn, B., Thomas, R., & Rawson, K. A. (2018). Learning more from feedback: Elaborating feedback with examples enhances concept learning. *Learning and Instruction*, *54*, 104-113.

Gurbanov, E. (2016). The Challenge of Grading in Self and Peer-Assessment (Undergraduate Students' and University Teachers' Perspectives). Online Submission, 1(2), 82-91.

Harris, L. R., & Brown, G. T. (2013). Opportunities and obstacles to consider when using peer-and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101-111.

Harris, L. R., Brown, G. T., & Harnett, J. A. (2015). Analysis of New Zealand primary and secondary student peer-and self-assessment comments: Applying Hattie and Timperley's feedback model. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 265-281.

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.

Ion, G., Sánchez Martí, A., & Agud Morell, I., (2019). Giving or receiving feedback: which is more beneficial to students' learning?. *Assessment & Evaluation in Higher Education, 44*(1), 124-138.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works?. *Change: the magazine of higher learning*, *30*(4), 26-35.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational psychology review*, *1*(4), 279-308.

Lambert, M. (2012). *A Beginner' s Guide to Doing Your Education Research Project*. Sage.

Lawrence, M. J. (1996). The effects of providing feedback on the characteristics of student responses to a videotaped high school physics assessment. *Unpublished doctoral thesis, Rutgers University, New Brunswick, NJ*.

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*(2), 193-211.

Li, L., Liu, X., & Steckelberg, A.L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525-536.

Liu, N. F., & Carless, D. (2006). Peer Feedback: The Learning Element of Peer Assessment. *Teaching in Higher Education 11*(3), 279-290.

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

Meusen-Beekman, K., & Joosten-ten Brinke, D. (2010). Peer assessment in het basisonderwijs. *EXAMENS, 3*, 13-16.

Ng, O. L., Ting, F., Lam, W. H., & Liu, M. (2020). Active learning in undergraduate mathematics tutorials via cooperative problem-based learning and peer assessment with interactive online whiteboards. *The Asia-Pacific Education Researcher*, *29*(3), 285-294.

Noonan, B. & Randy Duncan, C. (2005), Peer and Self-Assessment in High Schools. *Practical Assessment Research and Evaluation*. Volume 10, number 17.

O'Donnell, A. M., & Topping, K. (1998). Peers assessing peers: Possibilities and problems. *Peer-assisted learning*, 255-278.

Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In *Assessment for learning: Meeting the challenge of implementation* (pp. 311-326). Springer, Cham.

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, *44*(8), 1253-1278.

Pond, K., Ul-Haq, R., & Wade, W. (1995). Peer review: a precursor to peer assessment. *Innovations in Education and Training International*, 32(4), 314-323.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral science*, *28*(1), 4-13.

Rosenthal, R. and R.L. Rosnow (2008). *Essentials of Behavioral Research: Methods and Data Analysis, 3rd Edition*. New York: McGraw-Hill.

Rukavina, S., Zuvic-Butorac, M., Ledic, J., Milotic, B., & Jurdana-Sepic, R. (2012). Developing positive attitude towards science and mathematics through motivational classroom experiences. *Science education international*, *23*(1), 6-19.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin*, *143*(6), 565.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Stiggins, R. J. (2001). *Student-involved classroom assessment* 3rd edition. Upper Saddle River, New Jersey. Merrill-Prentice Hall.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British journal of educational studies*, *53*(4), 466-478.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, *68*(3), 249-276.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Springer, Dordrecht.

Topping, K. J. (2009). Peer assessment. *Theory into practice*, *48*(1), 20-27.

Topping, K. J. (2017). Peer assessment: learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, *1*(1), 1-17.

Topping, K. (2018). *Using Peer Assessment to Inspire Reflection and Learning*. Routledge.

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.

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal*, *31*(2), 369-395.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. West Palm Beach, FL: Learning Sciences International.

Warburton, K. (2003). Deep learning and education for sustainability. *International Journal of Sustainability in Higher Education*, *4*(1), 44-56.

Wu, Y., & Schunn, C. D. (2020). The Effects of Providing and Receiving Peer Feedback on Writing Performance and Learning of Secondary School Students. *American Educational Research Journal*, *20* (10), 1-35.

Websites:

RIVM (accessed 6 July 2020). Retrieved from: https://www.rivm.nl/en/novel-coronavirus-covid-19/in-depth-information

COVID-19 pandemic (accessed 6 July 2020). Retrieved from: https://en.wikipedia.org/wiki/COVID-19_pandemic

COVID-19 pandemic in the Netherlands, (accessed 6 July 2020). Retrieved from: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_the_Netherlands#March_2020

VO-Raad, (accessed 6 July 2020). Retrieved from (Dutch website): https://www.vo-raad.nl/nieuws/laatste-nieuws-rondom-coronavirus-en-onderwijs

Lesopafstand, (accessed 6 July 2020). Retrieved from (Dutch website): https://vo.lesopafstand.nl/lesopafstand/weer-op-school/

Appendix 1: Student guidelines for online reviewing and giving feedback

You will get homework to do after the explanation of the teacher. These are the steps you will take to do the homework and then review your group member's homework. Below you will find the steps you will take to complete the assignment for peer assessment (PA). On page 2 you will find the assignment PA and your homework.

Blue: to do on your own homework
Red: to do on the homework of your group

Doing homework

Step 1 Complete the homework in Word before the next lesson.
Step 2 Mail this Word file to the teacher. During the lesson, the answers to the exercises will be provided.
Step 3 Also send your homework to your designated group member before the next (online) class.

Check homework of group member

During the next lesson, you will review the homework and write down feedback for your group member. Time for this action, 10 minutes:

Step 4 Review your classmate's homework and write down on page 2 of the Word file what is right and wrong.
Step 5 Write down feedback about the homework (on page 2). As feedback, write down two stars (so, positive feedback) and a wish (so, what can be improved for next time).
Step 6 Mail this file to your group member.

Discuss homework

Next, discuss the reviewed homework and written feedback with your group member. Do you agree with the received feedback and reviewed work? What could be done differently? Time for this action, 5 minutes:

Step 7 Contact each other to discuss the reviewed homework and written feedback from each other. During the discussion, state arguments and reasons why this feedback and reviews was given. If you disagree with each other's arguments, try first to come to an agreement yourself before contacting the teacher.
Step 8 In the Self-reflection on page 2, write down what you have learned from the feedback and from the discussion in order to improve your work for the next time.
Step 9 Mail this completed file to your teacher.

Assignment and homework for June 12, 2020

This logo means that you first do the homework in your notebook and then paste a photo of it in this Word file.

Make assignments 24, 25, 26 and 28 of section 7.4

 Assignment 24	
 Assignment 25	
 Assignment 26	
 Assignment 28	

For each part of the homework, write whether it is right or wrong.

Assignment 24	
Assignment 25	
Assignment 26	
Assignment 28	

Write down feedback about your group member's homework.

Star 1	
Star 2	
Wish	

Self-reflection

Write here what you have learned from the checked assignment, the feedback given and from the discussion with your group member.

Appendix 2 Sorted students' feedback by feedback type (in Dutch)

Versterkende (*Reinforcing*) feedback	Didactisch (*Didactive*) feedback	Correctief (*Corrective*) feedback	Suggestief (*Suggestive*) feedback
- Netjes uitgeschreven. - Berekeningen waren goed. - Goede inzicht. - Netjes uitgewerkt. - Goed huiswerk gemaakt. - Ik zag dat je het snapte. - Alles goed. - Goede uitwerkingen. - Alles goed. - Alle stappen goed gedaan. - Veel goed. - Alles gemaakt. - Alles is goed! - Alles wat jij gemaakt hebt is goed. - Goede uitschrijvingen. - Goed gemaakt. - Je snapt de stof. - Goed gemaakt. - Je hebt de stof van 5 wel goed onder controle. - Je hebt het goed en snel gedaan. - Je hebt goed je best gedaan. - Je hebt de stof onder controle. - Veel gedaan in zo'n korte tijd. - Jij had maar 2 fout.	- Iets uitgebreider uitleggen waarom. - Beetje te slordig soms, effe goed nakijken. - Kijk goed welke punten je nodig hebt. - Nauwkeuriger opschrijven van de stappen.	- Let volgende keer op dat je goede getallen gebruikt. - Volgende keer opschrijven hoe je aan het antwoord komt.	- Meer berekeningen. - Netter schrijven. - Dat je dat goed op de foto hebt staan. - Alles goed hebben. - Huiswerk maken. - Beter je antwoorden nog eens nakijken. - Beter je antwoorden nakijken. - Je kan iets beter nakijken en controleren dat je alles goed hebt. - Iets beter thuis werk doen. - Iets netter schrijven. - Schrijf mooier, netter en overzichtelijker. - Je kan wat korter schrijven. - Bij een schets hoef je niet alles uitgebreid te schrijven. - Nakijken. - Goed uitwerken.

- Jij maakt de formule zo makkelijk mogelijk. - Je laat duidelijke berekeningen zien. - Het is goed te volgen. - Nette berekeningen. - Je snapt het hoe je moet berekenen. - Hartstikke mooi. - Je snapt de stof goed! - Geen slordigheidsfoutjes.			
Total: 31	4	2	15

Appendix 3

Figure 4: Message to the director of GSR in Rijswijk

Figure 5: WhatsApp messages between the math teacher and me concerning the permission of the director of GSR in Rijswijk

Appendix 4 Interview guide questions

Questionnaire questions:

Closed questions	Open questions
On what scale did you enjoy the peer assessment activity?	Following the feedback from your group member, how are you going to improve your way of learning? Please give a short explanation.
Did you learn anything from giving feedback?	What have you learned from the feedback you have received? Give one example you consider as the most important point.
Did you find PA useful?	What have you learned by giving feedback to your group member? Give one example you consider as the most important point.
To what extent do you agree with the feedback you received from your group member?	What went wrong with PA? Be critical.
Did the feedback contribute to your way of learning?	Please give a tip on how the PA activity could be improved.

Interview questions to students (meaty middle and rounding off):

- How was your experience with *reviewing* each other's work?
- What is your opinion about *giving feedback* to your group member?
- What is your opinion about *receiving feedback* from your group member?
- Were the students' guidelines in the homework assignment clear to you? Please specify your answer.
- How was the communication with your group member from home? Did you ask each other more questions about the assessed homework assignments? Please specify your answer.
- What is your opinion about distance learning?
- Do you have any tips for improving the PA-activities?

Interview questions for the teacher (meaty middle and rounding off):

Training PA:

- Do you think the given manual is clearly sufficient to implement PA in class?
- Was the time available for the training sufficient?
- What did you miss in this training about PA and what could have gone better?

PA as a formative form during homework assignment:

- What is your opinion about the PA assignment for reviewing homework?
- Do you think it contributed to the students learning to learn? Please specify your answer.
- What is your opinion about the students' participation in this research?
- Do you think the students enjoyed participating in this research? Where did you notice that?

PA in the future:

- Would you implement PA again in your mathematics class?
- Would you do it again as directed in the manual? If not, what would you do differently?
- Do you have any further tips for conducting this type of research in the future?