



CHALMERS
UNIVERSITY OF TECHNOLOGY



Exploring Variations in Frequent Passing Grade Tests in Mathematics

Upper secondary school students' perceptions

Master's thesis in Learning and Leadership

Rebecca Corcoran

DEPARTMENT OF COMMUNICATION AND LEARNING IN SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

www.chalmers.se

Exploring variations in frequent passing grade tests in mathematics

Upper secondary school students' perceptions

Rebecca Corcoran

© Rebecca Corcoran, 2025.

Supervisor: Linda Mattsson, Department of Mathematics & Natural Sciences, Blekinge
Institute of Technology

Co-supervisor: Samuel Bengmark, Mathematical Sciences, Chalmers University of Technology

Examiner: Philip Gerlee, Mathematical Sciences, Chalmers University of Technology

Master's Thesis 2025

Department of Communication and learning in science

Chalmers University of Technology

SE-412 96 Gothenburg

Acknowledgements

I would like to say a big Thank you to the teachers at the school, Lisa, Agnes, Leo, Sofie, Kristina, Bea and Mathias, for allowing me to carry out this study as well as helping and supporting me throughout this project. I want to give a special Thanks to Lisa Strand for meeting with me every week to discuss any ideas or questions linked to the project. I am also deeply grateful for my supervisors, Linda Mattsson and Samuel Bengmark, and would like to say Thank you for supporting me and helping me whenever I had questions throughout this project.

Abstract

Frequent Passing Grades tests (FPG) have been proven to be beneficial for enhancing students learning. In this study students at an upper secondary school were exposed to high-stakes and low-stakes FPG-tests as well as variations in follow-up work and timing for feedback with the purpose to find out which variations are more beneficial for increasing students' perceived learning. Students were then given two surveys to fill out with questions about their experiences. The responses from the surveys were compiled and analysed using thematic analysis. Additionally, data was collected about how many students passed the FPG-tests.

The result from the surveys indicates that students perceive they learn more from having high-stakes FPG-tests. Students express they retain the material for longer when having low-stakes FPG-tests as well as feeling less stressed. Students with lower grades feel they learn more from teacher-led assessment reviews whereas students with higher grades were more inclined to choose self-assessment as the follow-up work, they learn the most from. The relationship between type of follow-up work students feel they learn from, and their academic performance was found to be statistically significant. However, regarding the timing of feedback a statistically significant relationship between when students prefer feedback, and their grades could not be found. Although, students with higher grades seemed to prefer feedback immediately while students with lower grades preferred feedback to be delayed, this difference is likely due to random variation.

Keywords: frequent testing, mathematics test, high-stakes testing, low-stakes testing, second-chance testing, feedback.

Table of Contents

1. Introduction	1
1.1 Aim.....	2
1.2 Specification of question formulations	2
1.3 Limitations	2
2. Theory	4
2.1 High-stakes and Low-stakes testing.....	4
2.1.1 <i>High-stakes testing</i>	4
2.1.2 <i>Low-stakes testing</i>	5
2.2 Memory and testing effect.....	6
2.2.1 <i>Second-Chance Testing</i>	7
2.2.2 <i>Mastery Learning</i>	7
2.2.3 <i>Frequent testing</i>	8
2.3 Feedback in learning	8
2.3.1 <i>Teacher-Led Assessment Review</i>	9
2.3.2 <i>Self-Assessment</i>	10
2.3.3 <i>Peer-Assessment</i>	11
2.3.4 <i>Timing of Feedback</i>	12
3. Method	14
3.1 Survey.....	15
3.2 Data analysis	16
3.3 Sources of error	18
4. Results	19
4.1 Identified themes	19
4.1.1 <i>Learning</i>	20
4.1.2 <i>Performance</i>	21
4.1.3 <i>Feelings</i>	23
4.1.4 <i>Study habits</i>	24
4.2 High-stakes.....	25
4.2.1 <i>Follow-up work students feel they learn most from when having high-stakes FPG-tests</i>	26

4.2.2 Optimal timing for feedback on high-stakes FPG-tests.....	28
4.3 Low-stakes	29
4.3.1 Follow-up work students feel they learn best from when having low-stakes FPG-tests	30
4.3.2 Optimal timing for feedback on low-stakes FPG-tests	31
5. Discussion	33
6. Conclusion.....	37
References	39
Appendix	43

1. Introduction

Tests and exams have always held an important place in schools and education. Commonly, tests are used to measure students' knowledge and for providing students with grades. Standardised tests are also used as a way to rank students before they apply for universities (Roediger III et al., 2011). Besides being a way to measure how much students are learning in school, tests can also be used as a tool to help students learn. Studies have shown that students who are exposed to frequent tests throughout a course perform better at the end-of-course exam (Morphew et al., 2020). Tests also utilise something called the *testing effect*, which forces students to retrieve information from memory when being tested, resulting in their memory changing (Bjork & Ligon Bjork, 2025.). This in turn, results in students retaining the material for a longer period of time. The use of frequent tests in mathematics is something that has been evaluated at an upper secondary school located in Gothenburg, Sweden. The school have been implementing an alternative assessment form, called *Frequent Passing Grade tests* (FPG), in their mathematics courses to help further their student's education.

FPG-tests are given three times during each chapter throughout a mathematics course at the school. All FPG-tests are given before the chapter test, and students perform them digitally. The FPG-tests are shorter tests where students are given questions only at the grade level E. The Swedish grading system is based on the grades E to A, where E is the grade needed to pass a course. If students successfully pass all three FPG-tests they will be allowed to skip the E-questions on the chapter test, and as a result, are able to focus only on doing the C and A questions. If students fail the FPG-tests or miss one or more due to sickness, they are given a second chance to demonstrate their knowledge of the required abilities to reach the passing grade E during the chapter test instead. The chapter test is structured as any other exam, incorporating all grade levels and providing a reasonable time limit for each question. Students who have passed the FPG-tests are able to jump straight to the higher grade level questions, while the students who failed the FPG-tests are given a second chance to meet the requirements for the passing grade E. Using FPG-tests in this manner is high-stakes for students, since their scores might affect their grade. FPG-tests can also be low-stakes, providing students with an opportunity to assess their knowledge without it affecting their grades (Janison Assessments, 2022).

FPG-tests utilise the testing effect as a way to help students learn (Bjork & Ligon Bjork, 2025). By being given FPG-tests students are presented with the opportunity to practice their knowledge through testing, which helps them remember the material more effectively (Morphew et al., 2020). FPG-tests also offer a second chance for students to showcase their comprehension of the material, which has been exhibited to reduce stress and anxiety among students. This is called *second-chance testing*. Second-chance testing can also be seen as a form of mastery learning, since students are given a similar test, reinforcing their memory, helping them form a deeper understanding of the material (Emeka et al., 2023).

The school is mainly focused on offering programs in technology, IT, and design. The students who attend the school are generally high-achieving students with high grades. This results in the school's admission scores being rather high compared to other upper secondary schools in the area. The school have been utilising FPG-tests for a few years now, and have conducted previous studies about frequent passing grades recently (Svensson & Vesterberg, 2024). The school is now interested in further examining how it can embrace the benefits of different variations in FPG-tests and use them in the best way to support their students' education.

1.1 Aim

The aim of this thesis is to investigate whether high-stakes or low-stakes FPG-tests are more effective in enhancing students' achievement in mathematics, what types of follow-up activities students perceive as most supportive of their learning, and when students believe feedback should ideally be provided.

1.2 Specification of question formulations

1. Are high-stakes or low-stakes FPG-tests more effective for enhancing students' achievements in mathematics?
2. According to students' perceptions,
 - a. are high-stakes or low-stakes FPG-tests more effective for their learning?
 - b. what is the best way to perform follow-up work to support their learning?
 - c. should feedback be provided immediately or delayed, optimizing their learning?

1.3 Limitations

This study will focus on students' assessment results rather than their internal learning development. Additionally, this project will not be using a control group that aren't exposed to

FPG-tests as a comparison. The reason is that several prior studies already exist that demonstrate how second-chance testing, and the testing effect improve students' learning. Also, since this project will evaluate how to optimise the use of FPG-tests, a control group will not be needed. This study was also conducted only at one school which limits the generalisability of the results.

2. Theory

This section aims to provide a theoretical background for this thesis. Firstly, high-stakes and low-stakes testing will be presented, followed by different ways to utilise the testing effect to enhance students' learning. Furthermore, different types of follow-up work will be presented. The types of follow-up work that will be addressed are *teacher-led assessment reviews*, *self-assessment*, and *peer-assessment*. Finally, the optimal timing for feedback after assessments will be introduced.

2.1 High-stakes and Low-stakes testing

Several different assessment forms are utilised in schools today. Two of these types are high-stakes and low-stakes tests (Janison Assessments, 2022). Tests are an example of a summative assessment procedure. A summative assessment is a method of assessment where students' knowledge are summarised at a given point in time to see if they have reached the learning objectives of the course (Bates, 2023). Tests are one example used for summative assessments; however, discussions or essays during a course can also be used as summative assessments. A formative assessment on the other hand, takes place throughout the course and makes sure that students are learning during the year. Formative assessments actively involve learners in their own learning process by providing regular feedback (Bates, 2023). FPG-tests can be considered as a summative assessment form that can also be used in a formative way to promote student learning.

The school at which this study takes place have previously been utilising high-stakes FPG-tests, meaning they might affect students' grades. Low-stakes FPG-tests are something the school hasn't attempted before but is interested in doing so now.

2.1.1 High-stakes testing

High-stakes tests are defined as assessments where the outcome has a significant impact on students' futures (Janison Assessments, 2022). High-stakes tests can also sometimes be called standardised tests and are in that case commonly used to determine a student's final grades in different courses or as an entrance exam for university studies (Demir & Keleş, 2021). High-stakes tests can also be used to either fail or pass students in a course (Demir & Keleş, 2021). At the school in this study, high-stakes FPG-tests are used for allowing students to skip E questions during the chapter test if they passed them on the FPG-tests.

Studies have displayed that high-stakes tests result in students studying more continuously meaning they stay on top of their readings (Demir & Keleş, 2021). This, in turn, gives rise to students performing better on tests as well as learning how to study more effectively. However, studies have also revealed that students' anxiety and stress levels increase when they are given high-stakes tests (Demir & Keleş, 2021). High-stakes tests can also result in a decrease in students' intrinsic motivation, which in turn proposes the risk that students become passive in their learning process (Demir & Keleş, 2021). This is due to the possibility that students only memorise information in order to pass the high-stakes test instead of actually learning the material long term (Demir & Keleş, 2021). High-stakes tests have also revealed to shift teaching toward a more test-oriented education, where teachers mainly focus on what is being tested, potentially resulting in parts of the syllabus being ignored. This could be contributed by the pressure many teachers experience when having to give out high-stakes tests.

The impact of high-stakes tests on learning and teaching is referred to as *washback*. In a broad sense, washback can be defined as the extent to which a test affects school policies, parents, and school administrators. At a more specific level, washback can be defined as how the test affects students' ability to learn as well as the teacher's ability to teach (Demir & Keleş, 2021). The term washback can also be defined as something teachers and students would not normally engage in but undertake due to the presence of a test (Demir & Keleş, 2021). This means that tests affect what teachers cover in lectures and teaches their students. A consequence of this could also result in education being more teacher-centred, focusing on some parts of the course syllabus and missing other parts, and in that way limit students' learning (Demir & Keleş, 2021). However, washback can be present in varying levels and can therefore be seen as both positive and negative (Barnes, 2017).

2.1.2 *Low-stakes testing*

Low-stakes tests are tailored to support students' learning. It is used as a formative assessment tool focusing on the learning process rather than the results (Schüttpelz-Brauns et al., 2020). The outcome of low-stakes tests does not impact the future of the students and is instead used as a way to help students identify learning gaps (Janison Assessments, 2022). Therefore, low-stakes assessments have an important function to help students develop self-regulated learning (Schüttpelz-Brauns et al., 2020). Low-stakes tests can be used as a way to show students their learning progress in a subject. This also provides the students with information on what they need to focus on in order to reach a higher grade.

Studies demonstrate that low-stakes tests help reduce stress and anxiety for students (Simzar et al., 2015). However, a consequence that arises regarding low-stakes assessments is lower motivation in students (Simzar et al., 2015). If students don't feel motivated to perform and demonstrate their knowledge on a low-stakes assessment, their results decrease. This, in return, leads to the risk of students not being able to fully showcase their abilities, and therefore might not be able to entirely learn from the tests (Penk et al., 2014). However, when students partake in a low-stakes assessment, the *testing effect* is utilised which leads to students learning during the test (Walck-Shannon et al., 2019). Thus, an increase in students' performance on their final exams has been noted when students have been exposed to frequent low-stakes tests throughout a course (Walck-Shannon et al., 2019).

2.2 Memory and testing effect

There are three different stages of memory: *encoding*, *retention*, and *retrieval*. Encoding is the process through which our brain transforms an experience into something our brain can store, typically through visual, acoustic or semantic means. Retention is where the information is kept as a memory and retrieval is when the memory is successfully accessed (East Carolina University, 2024).

Memory can also be divided into *sensory memory*, *short-term memory* and *long-term memory*. Sensory memory is where information is stored during a very short time span, between half a second up to four seconds. Short-term memory is also known as active or working memory, which is where information that we are actively thinking about is stored. Information in our short-term memory can be stored for 20-30 seconds. Long-term memory is where information is stored continuously. Information stored in long-term memory can be retrieved and accessed when needed. Long-term memory is also known as the unconscious memory (East Carolina University, 2024).

By connecting learning to memory and the ability to retrieve information from our long-term memory, students' performance has been demonstrated to increase. Taking tests in school helps reinforce the storage of information in our long-term memory, a phenomenon known as the *testing effect* (Bjork & Ligon Bjork, n.d.). When students are tested on their knowledge and successfully retrieve information, their memory will change, and their ability to learn will

increase. Research has shown that applying the testing effect in classrooms improves students' retention and, consequently, their academic performance. The testing effect is beneficial for students as it gives them the opportunity to practice retrieving information which leads to deeper encoding, storing the information in the long-term memory (Soderstrom & Bjork, 2023).

2.2.1 Second-Chance Testing

Second-chance testing is an approach of testing where students are allowed to do a similar assessment twice (Emeka et al., 2023). Second-chance testing also utilises the testing effect which has been shown to have a positive impact on students' retention and therefore their academic results (Emeka et al., 2023). Studies have also demonstrated that second-chance testing reduces students' stress and anxiety levels. Since second-chance testing permits students to retake a similar test they have already taken, their grades have been revealed to increase (Emeka et al., 2023). Second-chance testing has also been known to reduce the number of students who fail and drop out.

The FPG-tests at the school always provide students with a second chance to show their knowledge, regardless of whether the FPG-tests are high-stakes or low-stakes. If the students have high-stakes FPG-tests and fail, they are given another opportunity to reach the passing grade E. This is true for low-stakes FPG-tests as well, only the difference being that all students need to do the E questions twice, knowing from the start they have two times to show that they comprehend the material.

2.2.2 Mastery Learning

Mastery learning allows the learner to master one level before moving on to the next (Bates, 2023). Students are introduced to a concept and then assessed in some way, either by a quiz or a test, before moving on to the next concept. Students are allowed to correct their mistakes which improves their memory before they tackle the next level. Several studies have shown that mastery learning positively affects students' learning. However, applying a mastery learning model comes with challenges. One challenge that arises in regard to mastery learning is that it takes much time both for students and teachers (Bouchrika, 2025). Teachers need to ensure all students have mastered a subject before moving on, but they must also provide for the group. Therefore, teachers need to make individual study plans for their students which takes time and resources from the teacher (Bouchrika, 2025).

FPG-tests at the school can be viewed as a sort of mastery learning, since students either need to complete the E questions twice if they have low-stakes FPG-tests or since they have to show that they master the questions on the grade level E on the high-stakes FPG-tests before moving on to the next grade level, and thus the next concept. This increases their chance of learning the material as they are allowed to first only focus on the E level before moving on to higher grade levels during the chapter test.

2.2.3 Frequent testing

Frequent testing refers to having an increased number of tests during a course to help students learn. The purpose of having frequent tests is not to introduce students to new concepts but to help them remember what they have already learned (Morphew et al., 2020). Frequent testing utilises the testing effect which leads to improved retention among students. Studies have also demonstrated that students who were exposed to frequent testing performed better during their final exam (Morphew et al., 2020). Frequent testing has also been reported to have a positive effect on students' study habits. When exposed to frequent tests or quizzes student are more likely to keep up with their lectures and readings (University of Texas at Austin, 2025). Frequent testing does not have to be limited to tests. It can also be questions for students during a lecture, a quiz, flashcards, or a brain dump (Muzsnay et al., 2024). Essentially, testing is everything that forces students to rely on their memory without the help of external sources (Muzsnay et al., 2024).

Since students need to partake in three FPG-tests they are frequently tested on the material before they take the chapter test. FPG-tests are therefore a way for students to regularly check their knowledge on the material as well as enhancing their memory through the testing effect, which require students to actively recall information.

2.3 Feedback in learning

Studies have been made that demonstrate how feedback can be used to enhance students' learning (Bates, 2023). To give feedback, there must be motive, opportunity, and means. Motive refers to the fact that there must be a will to give feedback. An opportunity where feedback can be given must also be present alongside means of using feedback effectively. There are different points that need to be considered when exploiting feedback to enhance learning. It is important that the feedback focuses on the task and not the person. The feedback must be elaborated, and it should be presented in reasonable amounts to avoid making the person receiving the feedback

feel overwhelmed and attacked. Feedback should be presented in a simple and clear way for it to be most effective. Furthermore, it is also important to be objective when giving feedback as well as the feedback promoting the relevant learning objectives (Bates, 2023).

Feedback can work as a tool for students to identify where they are right now and what they need to do to reach the next step of progress (Hattie & Timperley, 2007). Feedback can help students develop their self-efficacy and therefore gain confidence in their own learning and understanding of the material (Bates, 2023). This can also further help students become motivated. However, it is important to be aware that feedback can have a negative effect on students' motivation and therefore it is important that the students know that it is okay to fail at a task and learn from it (Bates, 2023).

Feedback can be divided into *formative feedback* and *summative feedback*. Formative feedback can be defined as a process that takes place during the entirety of a course. It is used as a way to make sure that students are learning and keeping up with the material by being given feedback consistently (Bates, 2023). A way to work with formative feedback is to use self-assessment where students assess themselves and their performance. Summative feedback, on the other hand, is given at a specific time during the course. This could be feedback at the end of a lesson or a discussion. A common sort of summative assessment is a test where students are given points based on their performance, which are then calculated toward a final grade (Bates, 2023).

In this study, the effect on students' learning in relation to FPG-tests of three different types of follow-up work is being evaluated. These three are *teacher-led assessment review*, *self-assessment*, and *peer-assessment*. The timing for feedback and assessment work is also being evaluated in this study, that is, when is the best time to give students feedback to optimise their learning.

2.3.1 *Teacher-Led Assessment Review*

Teacher-led assessment reviews are a well-established method used in the teaching world. Teacher reviews after examinations are used to help students see common mistakes made during tests. It has been demonstrated that correcting errors can help students learn and improve their memory, helping them recall the correct answer for a longer period of time (Kemp, 2021). However, it takes more than just stating what the correct and incorrect answers are.

Additionally, during a teacher review, students are mostly passive, which can affect how much they learn. For teacher-led assessment reviews to be effective, students must go from being passive learners to active learners. Teachers showing students how to reach the correct answer and engaging them in the solution process is more effective for student learning than just hearing the right answer (Kemp, 2021). If the teacher-led test walkthrough is to be effective, teachers need to engage students and give them room to think critically about the material. There is also a risk that teacher-led reviews mostly focus on the class as a whole and fail to address the individual challenges that students face (Favero & Hendricks, 2016).

Metacognition and self-efficacy play an important part in how students learn (Cera et al., 2013). High-achieving students often have a higher self-efficacy and better understanding of what they need to do to learn. Low-achieving students have a reduced metacognition and understanding of how they learn (Kemp, 2021). Although, teacher-led correction reviews have been seen to promote low-achieving students' use of metacognition and therefore enhancing their learning process, high-achieving students already possess this skill. Studies also display that teacher-led assessment reviews are beneficial when students encounter similar problems in the future, which is only the case if the course material is cumulative (Favero & Hendricks, 2016).

Studies also demonstrate that teacher-led assessment reviews increase the dialogue between students and teachers. This makes it easier for the teacher to realise if students have grasped a concept or if they need further instructions (Kemp, 2021). Teacher-led assessment reviews also provide teachers with an opportunity to reflect on their own teaching and improve their instruction. Furthermore, teacher-led test walkthroughs enable summative tests to be used in a formative manner, helping students reflect on their learning. One risk that arises in regard to teacher-led assessment reviews is that students might feel competitive with each other and compare themselves to their peers which can cause a negative classroom environment (Kemp, 2021).

2.3.2 Self-Assessment

Self-assessment is a way to assess and critique one's work. During self-assessments, the student will reflect on their performance and, as a consequence, become more involved in their learning progress (Panadero et al., 2017). Assessing one's work helps students regulate their learning since it forces them to be an active part of their own learning progress. Self-assessment helps students understand what is expected of them and what they need to do to reach a certain criteria

(Kangaslampi et al., 2022). Alongside this, self-assessment helps students gain insight into how accurate and reasonable their answers are, furthering their learning process (Panadero et al., 2017). Studies have also indicated that when students self-assess their performance they are more interested in their work, and their motivation increases since students are given more autonomy (Panadero et al., 2017).

One problem that may arise with self-assessment is that high-achieving students are more prone to underestimate their performance, whereas low-achieving students tend to overestimate their performance (Kangaslampi et al., 2022). Students who struggle to understand the material may also have a harder time self-assessing their work. Likewise, it has been revealed that self-assessment is not as effective unless the student understands the material well enough (Kangaslampi et al., 2022).

2.3.3 Peer-Assessment

Peer-assessment is a form of assessment where students evaluate each other's work (Biton, 2025). Peer-assessment promotes interaction among students which has been shown to increase motivation and engagement. By being exposed to different perspectives through peers, students' understanding of the material has been demonstrated to increase (Biton, 2025).

Since peer-assessment involves giving and receiving feedback, it is important that students understand the material and the task at hand to be able to give and receive proper feedback (Biton, 2025). This provides students with the opportunity to practice their critical thinking and reflective skills. Another benefit of peer-assessment is that students gain another responsibility for their learning since they need to understand how their peers think to be able to give them feedback. This also forces students to see the learning task from the teacher's perspective since they are assessing each other and, therefore, need to understand the goals of the assessment.

Although peer-assessment presents several benefits, there are also challenges associated with this type of feedback. Students need to trust each other. Otherwise, they will experience problems with utilising the feedback they have been given. Peer-assessment can also lead to anxiety and stress among students if they feel that they are being judged by their peers (Biton, 2025). It is therefore important that the teacher presents clear instructions for students on how they will perform peer-assessment as well as letting them practice (Biton, 2025).

Implementing a structure for peer-assessments in the classroom takes time and effort. To facilitate the implementation of peer-assessments in the classroom teachers can follow three steps: *develop*, *establish*, and *enhance* (Bates, 2023, p. 316-317). Develop refers to teachers implementing strategies for peer-assessment as well as processes for how the assessment will be performed. During the establishment, the teacher provides students with an outline of the success criteria, which allows them to help decide on the processes they prefer. The last stage is enhance, which means that teachers work together with students to identify suitable success criteria for the planned learning outcomes (Bates, 2023, p. 316-317).

Peer-assessment can be anything from checking answers against goals to discussing and comparing work among peers (Bates, 2023, p. 316-317). Peer-assessment also allows students to gain a certain amount of power over their learning, since they are involved in assessing each other. Students can also access some feedback that the teacher cannot reach since students see each other as peers. Students speak the same language, which makes it easier for them to listen to and trust their peers compared to their teachers (Bates, 2023, 316-317).

2.3.4 Timing of Feedback

There have been several previous studies about the effect of immediate feedback or delayed feedback on students' learning. However, the definitions of immediate and delayed feedback differ. Previous studies have defined immediate feedback as feedback given immediately after completing a test or a set of assignments (Smith et al., 2022). Delayed feedback, however, can be defined as feedback given anytime from minutes after a test up to days after (Smith et al., 2022). Although many seem to agree that delayed feedback implies feedback given somewhere between 24-48 hours after an assessment (Brosvic & Epstein, 2007). Immediate feedback has been shown to have a positive effect on learners' retention ability (Brosvic & Epstein, 2007). Studies have also indicated that the timing of feedback should be adjusted depending on the learners. Furthermore, studies reveals that high-achieving learners should be given delayed feedback, while low-achieving students should be given feedback immediately after a task (Bates, 2023).

The idea of immediate feedback derives from the early ideas of behaviourism (Mullet et al., 2014). Behaviourism is based on the idea of conditioning and receiving either a reward or a punishment. It was clear that when feedback was given immediately in the form of a punishment or reward improved results were reached (Mullet et al., 2014). Many still consider this to be

important when advocating for immediate feedback. During the rise of cognitivism feedback was changed to be viewed as information to the learner rather than something mechanical (Mullet et al., 2014). During this era, it was argued that delayed feedback had a higher positive impact on the learners' long-term memory and retention (Mullet et al., 2014).

A part of this study is to examine what students feel is the optimal timing to be given feedback for their learning regarding FPG-tests. Feedback on FPG-tests can be given immediately, i.e. on the same lesson as the FPG-tests, or it can be delayed until the lesson after the FPG-test was given. Feedback provided on the FPG-tests is given either in the form of teacher-led assessment reviews by the teacher, self-assessment from students or peer-assessment among students.

3. Method

The data in this study consisted of two student surveys. Six different classes, a total of 170 students from the same upper secondary school, were asked to take part in this research. Each class was given around 10-20 minutes to complete the surveys during their mandatory math lesson. The students who participated in this research were enrolled in the technology program and were either in year 1 or year 2. Hereafter, the classes will be referred to as classes Alfa, Beta, Gamma, Delta, Epsilon and Zeta.

The study took place over the course of 12 weeks where students were exposed to different types and variations of FPG-tests. The study was divided into two six-week phases. During the first six weeks classes Alfa, Beta, and Gamma were exposed to high-stakes FPG-tests which means that they had the chance to reach the passing grade E. Classes Delta, Epsilon, and Zeta were exposed to low-stakes FPG-tests during the first six weeks, where the FPG-tests were used for reference only, giving students an indication of what their strengths and weaknesses are. Alongside this, three different types of follow-up work were evaluated. These were teacher-led assessment review, self-assessment, and peer-assessment. All classes were exposed to all kinds of follow-up work. Lastly, the timing for feedback was also investigated, and students received feedback either immediately after the FPG-tests or feedback was delayed until the next lesson. If students were given a teacher-led assessment review immediately after a FPG-test during the first six weeks, teachers would then delay the teacher-led assessment review during the following six weeks. After the initial six-weeks phase the first survey was distributed to students, and then the classes switched from high-stakes to low-stakes and vice versa. After the following six-weeks phase the second survey was handed out to students.

The reason for distributing the survey twice was to account for the fact that, during the first six weeks, half of the classes were exposed to high-stakes FPG-tests while the other half undertook low-stakes FPG-tests. Repeating the same survey on two occasions helped maintain students' experiences for more consistent responses.

A summary of the different variations of FPG-tests that the six classes were exposed to can be found in Figure 1. Each class was given three FPG-tests during each six-weeks, followed by a chapter test covering the area they have been working on. In total students had sat through six FPG-tests each and two chapter tests at the end of the 12 weeks. The teachers at the school

documented the different variations and combinations of FPG-tests that they performed throughout the study.

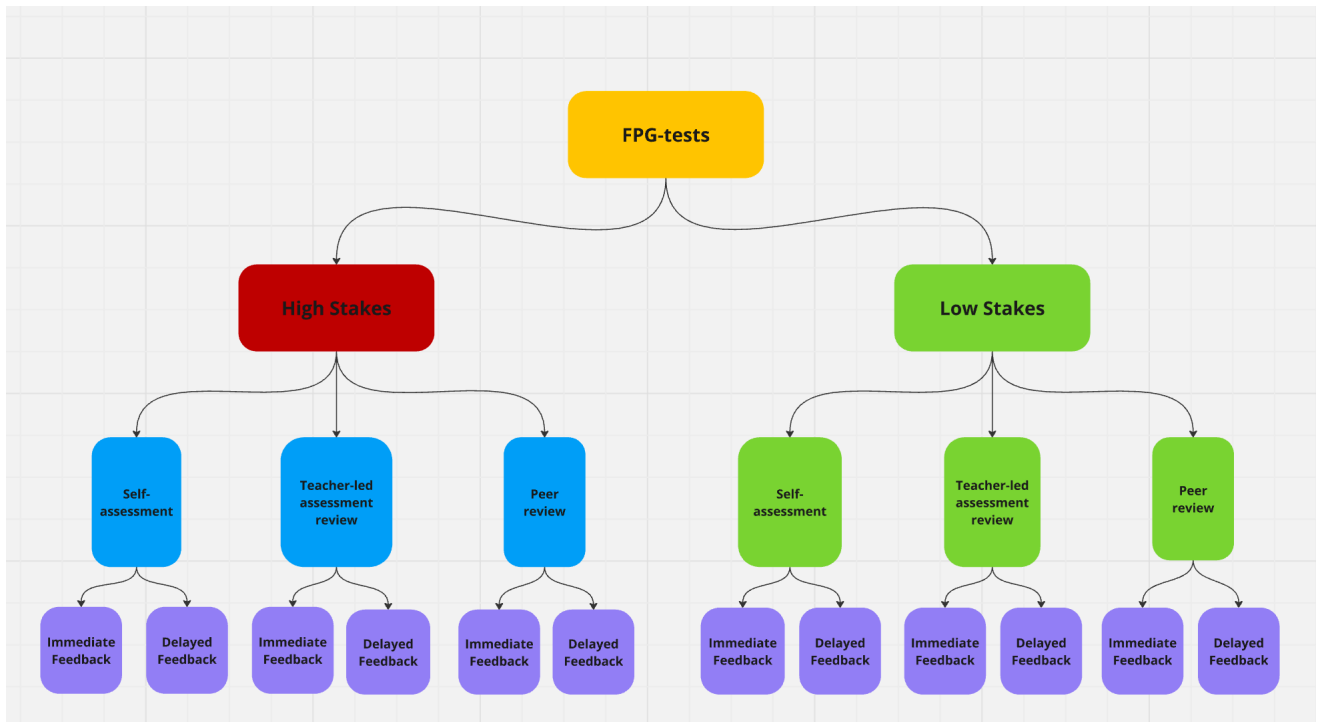


Figure 1: Overview of the different variations and combinations of FPG-tests that the students were exposed to.

3.1 Survey

The surveys were created together with the teachers at the school along with one other researcher. The survey was constructed based on a similar survey carried out at the school a year earlier. The questions were, however, changed to fit the current aim of this study. Due to the fact that a similar study had already been conducted at the school the survey questions in this project were not piloted. The topics of interest in the survey included what type of FPG-tests students think impact their learning the most, the type of follow-up work that best supports their understanding, and the optimal timing for feedback.

The survey consisted of three main closed-ended questions regarding type of FPG-test, follow-up work and timing for feedback. These were followed by two open-ended questions where students were able to elaborate on their responses. The first open-ended question allowed students to address their opinions on high-stakes and low-stakes FPG-tests. The second open-ended question allowed students to express why they thought they learned more from a certain type of follow-up work and a specific time for feedback. The survey questions are attached in Appendix 1.

The students were given a consent form to fill out before answering the survey. If they agreed to be a part of the study, they then got to answer the survey during their mandatory math class. Students were given the same survey twice, first after six weeks when they had only been exposed to one branch of FPG-tests and the different variations of follow-up work and timing for feedback. Then again after another six weeks when they had switched and been exposed to both branches and all different variations of FPG-tests. The responses to the survey were pseudo-anonymous. Students received a unique code to enter the survey while still maintaining anonymity in the survey responses.

3.2 Data analysis

The responses from the surveys were compiled and analysed at two separate points in time. Once both surveys had been analysed separately their findings were reviewed in order to identify the most frequently valued elements (Bjorkman, 2015).

The responses to the quantitative questions from both surveys were categorised into whether if students had been exposed to high-stakes or low-stakes FPG-tests at the time of their response. The replies to the first open-ended questions from both surveys were merged and sorted into groups based on similarity in responses (Esaiasson et al., 2017). From these responses, four themes were identified: learning, performance, feelings and study habits. The theme of learning encompasses students' experience with how they use the FPG-tests to learn or check that they are in fact learning. Performance highlights students' perception of how FPG-tests affect their performance. In this theme the category *second chance* could also be identified, referring to students having several chances to reach a certain grade. Feelings refer to students' emotions in relation to FPG-tests. The theme of feelings was divided into *positive feelings* and *negative feelings*. Positive feelings are feelings that students experience that make them happy or more confident. Positive feelings also include decreased stress and anxiety. Negative feelings are feelings associated with an experience of increased stress, pressure or anxiety in connection with FPG-tests. The last theme that was found was study habits. Study habits illustrate how FPG-tests affect students' study routines as well as motivation. Some responses from students may fit two themes. Table 1 displays the four themes and the number of items found for each theme.

Number of items	Themes
80	Learning
201	Performance
113	Feelings
74	Study habits

Table 1: Number of items found for each theme in both surveys.

Table 2 shows three examples of how a few of students' responses to the first open-ended question were sorted into the four themes. The first response was sorted into the theme of learning since it highlights that low-stakes FPG-tests help students retain the information for a longer period of time. The second response was sorted into performance since it has to do with being able to reach the grade E. Lastly, the third response was sorted into feelings since it deals with not feeling stressed about having to do the E questions during the chapter test.

Original response	Translated response	Theme it fit into
<i>“Fördelen med endast avstämning skulle jag säga är att du behåller informationen längre”</i>	<i>“The advantage with reference only I would say is that you keep the information for longer”</i>	Learning
<i>“Man har ett garanterat E, vilket ger en mer tid till de svårare uppgifterna. Det finns inga nackdelar överhuvudtaget.”</i>	<i>“You are guaranteed an E, which gives you more time for the harder questions. There are no disadvantages at all”</i>	Performance
<i>“Behöver inte stressa över e del inför provet”</i>	<i>“Don't need to stress over E questions before the test”</i>	Feelings

Table 2: Examples of students' responses and how they were categorised.

The data gathered from the second open-ended question, concerning why students feel they learn more through a specific type of follow-up work or when they feel the optimal timing for feedback is, were too limited for meaningful analysis. The majority of students simply responded with the same answer they had provided in the closed-ended questions. As a result,

these responses were too sparse to draw any conclusions from and were therefore excluded from further analysis.

To analyse whether a statistically significant relationship could be found between the type of follow-up work students feel they learn the most from and what grades they have a chi-squared test was performed (Zaveri, 2024). The null hypothesis that was tested was that no significant relationship could be found between students' grades and their preferred type of follow-up work. The alternative hypothesis was that a statistically significant relationship could be identified. A chi-squared test was also performed to examine if a statistically significant relationship could be found between when students think the optimal timing for feedback is and their grades. The null hypothesis was that no significance could be found, and the alternative hypothesis was that a statistically significant relationship was present.

3.3 Sources of error

Sources of error that could affect the reliability of the results obtained in this study are presented in this section.

One aspect that could affect the reliability of the results is the fact that one class, class Epsilon, had a substitute teacher when they were supposed to do the high-stakes FPG-tests. This substitute teacher was under the impression that the FPG-tests were optional for students, which resulted in only two students undertaking them. As a consequence, the results from these FPG-tests cannot be used to draw any conclusions from. However, since this was only the case when class Epsilon were exposed to high-stakes FPG-tests, the class will not be excluded completely.

Another aspect that can affect the reliability of the results is the fact that students aren't allowed to redo FPG-tests if they miss one. If students miss one or more FPG-tests due to sickness they will automatically be marked as failed, even though the student might not have failed if they had done the FPG-test. This is something to keep in mind, although it wasn't taken into consideration when analysing the results.

4. Results

In this chapter the results from the surveys will be presented. The responses to the open-ended question from both surveys were combined and are presented according to the themes that emerged. For each theme, examples of students' responses are included along with if the student had high-stakes or low-stakes FPG-tests at the time for their given response. The responses to the closed-ended questions were categorised based on whether students had been exposed to low-stakes or high-stakes FPG-tests. These quantitative results will be presented separately for each group.

A total of 157 students responded to the first survey and 139 responded to the second survey. Figure 2 shows the percentages of students with each grade for year 1 and year 2. In year 1, a total of 76 students responded to the surveys, which corresponds to 48,1%. In year 2, a total of 82 students participated in the survey, which corresponds to 51,9%. There were slightly more students in year 2 who responded to the survey, but not a big difference.

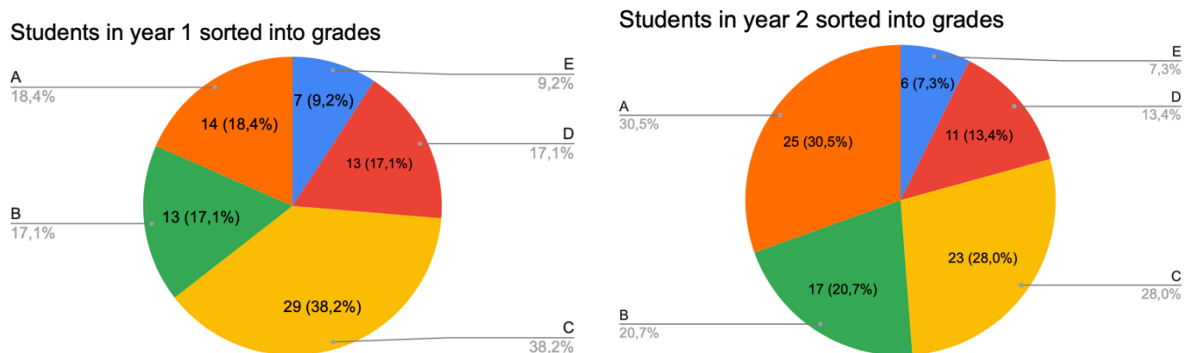


Figure 2: Number of students in each year sorted into grades who participated in the study.

4.1 Identified themes

The responses from both surveys were sorted into the four identified themes. Table 3 shows the number of items found for each theme. Some responses may fit in two or more themes.

Number of items	Themes
80	Learning
201	Performance
113	Feelings
74	Study habits

Table 3: Number of items found for each theme.

Below, the findings of each theme will be presented along with chosen responses from students that represent the group as a whole including what type of FPG-tests the student had at the time they gave their response.

4.1.1 Learning

From both surveys a total of 80 responses were categorised as ways students learn. Several students highlight that FPG-tests can be used to check that you understand the material. This was listed as a positive mostly when the FPG-tests were given as low-stakes. These responses also include that low-stakes FPG-tests help them remember the material for longer since they need to do the E questions on the chapter test as well as during the FPG-tests. In Table 4 two responses can be seen as an example of students' perspective on learning. Students express the value of being able to check if they have learnt the material regardless of if they have high-stakes FPG-tests or low-stakes FPG-test.

Original response	Translated response	Type of FPG-tests
<i>“jag tycker att det är bra att få en kort avstämning på det vi har lärt oss så att man kan vara säker på att man har hängt med.”</i>	<i>“I think it is good to get a short checkpoint of what we have learnt so that you can be sure that you are following along”</i>	High-stakes
<i>“Fördelen med det är att man kan kolla om man verkligen kan grunderna i arbetsområdet man jobbar med. Men man får även med sig kunskaperna under längre tid t.ex istället för plugga bara på det som är E nivå och använder det bara på E-testet kommer man behöva det längre fram på provet [...].”</i>	<i>“The pro with it is that you can check if you really know the basics in the chapter you are working with. But you also bring the knowledge during a longer time for example than if you only studied for what is at E level and use it only on the FPG-test you will need it on the chapter test further along”</i>	Low-stakes

Table 4: Two quotes that represent students experience of how the FPG-tests contribute to their learning.

A disadvantage that students bring to light in connection to high-stakes FPG-tests is the fact that the basic knowledge isn't repeated since they don't complete the E questions on the chapter tests if they pass the FPG-tests. However, when they are given low-stakes FPG-tests they repeat the material numerous times. Several students also express that high-stakes FPG-tests leads to them only focusing on E level questions in order to pass the FPG-tests and then don't revise the

fundamentals before the chapter test. Table 5 highlights two responses from students about forgetting the material before the chapter test that are representative of the whole group.

Original response	Translated response	Type of FPG-tests
<i>“Nackdelen blir väl att de "enklare" sakerna inte blir repeterade [...]”</i>	<i>“The con is probably that the “easy” stuff isn’t repeated [...]”</i>	Low-stakes
<i>“[...] nackdel är att man kanske glömmer av saker innan prover som annars kanske hade förekommit på t.ex övningsprov eller liknande”</i>	<i>“[...] a con is that you maybe forget the things before the test that otherwise would have been on a practice test or similar”</i>	High-stakes

Table 5: Responses from students in connection to learning and memory.

Furthermore, Table 6 displays that students experience that they remember the material for longer when having low-stakes FPG-tests as well as being given more time to work during the lesson.

Original response	Translated response	Type of FPG-tests
<i>“Fördelen med endast avstämning skulle jag säga är att du behåller informationen längre”</i>	<i>“The advantage of having low-stakes is I would say that I keep the information for longer”</i>	High-stakes
<i>“[...] Det positiva är dock att man får mer tid att jobba på lektionen.”</i>	<i>“[...] The positive is that you get more time to work during the lesson”</i>	Low-stakes

Table 6: Responses from students regarding learning and memory.

4.1.2 Performance

In total, from both surveys, 201 responses were identified as relating to performance. The majority of responses were about saving time on the chapter test by skipping the E questions and therefore being able to focus on achieving a higher grade. A few students express that high-stakes FPG-tests leads to making less careless mistakes. This is because they save time during the chapter test and don’t have to stress through all the questions, leading to fewer careless mistakes being made. Table 7 show two responses from students saying they can focus on C and A questions when they don’t have to do the E questions as well, which is representative of

the group as a whole. This is something students experience as important even when they had low-stakes FPG-tests.

Original response	Translated response	Type of FPG-tests
<i>“Fördel med E-test är att man har möjligheten till att klara av provet sen fokusera på att nå högre nivåer. [...]”</i>	<i>“Advantages with FPG-test is that you have the opportunity to pass the test and then focus on reaching higher levels [...]”</i>	Low-stakes
<i>“Att man kan lägga mer fokus och tid på de svårare frågorna, och jag ser inga konkreta nackdelar med det”</i>	<i>“That you can put more focus and time on the harder questions, and I don't see any concrete disadvantages with that”</i>	High-stakes

Table 7: Responses from students regarding performance and saving time on the chapter test.

Students also highlights the importance of being able to reach the passing grade E. Students feel that high-stakes FPG-tests play an important role in making sure they can reach all the criteria needed for the grade E. Students who commented this also think FPG-tests should be high-stakes since they then get confirmation from the teacher that they can pass the course. Table 8 contains two quotes from students highlighting the importance of getting validation that they can pass the criteria for the grade E. This confirmation from the teacher isn't provided for students when they have low-stakes FPG-tests.

Original response	Translated response	Type of FPG-tests
<i>“För att veta om man klarar E nivå eller inte”</i>	<i>“To know if you can do E level or not”</i>	Low-stakes
<i>“Fördelen är att få en bekräftad godkänn nivå.”</i>	<i>“The advantage is to get confirmed at passing level”</i>	Low-stakes

Table 8: Responses from students highlighting the importance of being able to reach the grade E.

A few students also express the importance of being given a second chance to show their knowledge. If they miss or fail one of the FPG-tests they know they have another chance to show that they actually comprehend the material during the chapter test. This leads to students feeling less stressed about having a bad day on the day of the chapter test. Table 9 shows two responses that encompass what students had to say about being given a second chance to reach the grade level E.

Original response	Translated response	Type of FPG-tests
<i>“plus klara nivån E, mer chans för att få minst E”</i>	<i>“pro pass E, more chances to reach at least E”</i>	Low-stakes
<i>“Jag anser att det inte finns några nackdelar med att inte för chans till avklarad E-nivå. Fördelarna är dock många; till exempel blir man mindre stressad på provdagen, har flera chanser att visa att man kan E-nivå och har mer till till frågorna på högre nivåer.”</i>	<i>“I consider that there are no cons with not having chance for completed E-level. The advantages are many; for example you get less stressed on the test day, have more chances to show that you can E-level and have more time to for the questions on the higher levels”</i>	High-stakes

Table 9: Responses from students touching on second chances.

4.1.3 Feelings

There were 113 responses identified as feelings. These could be categorised as positive feelings and as negative feelings. Among the responses categorised as positive feelings students named decreased stress levels and more confidence most frequently. Students also highlight a feeling of relief when having high-stakes FPG-tests and being able to walk into the chapter test knowing they already have reached a passing grade in the course. In Table 10 two responses have been chosen that show the positive feelings students experience. The second response also highlights that it can be stressful to have several FPG-tests, however the feeling of going into the test knowing you already have passed outweighs that stress. These students have both had high-stakes FPG-tests.

Original response	Translated response	Type of FPG-tests
<i>“Fördel: mer självförtroende”</i>	<i>“Pro: more confidence”</i>	High-stakes
<i>“Det kanske kan kännas stressigt att ha små prov men enligt mig är det jätteskönt att veta att man inte kommer få underkänt även fast det går dåligt på ett prov.”</i>	<i>“It can maybe feel stressful to have many small tests but in my opinion it’s really nice to know you can’t fail even if it goes bad on a test”</i>	High-stakes

Table 10: Responses from students representing the views on positive feelings in connection to FPG-tests.

This is also a recurring theme in the responses categorised as negative feelings. Students express that they feel more stress when having high-stakes FPG-tests, but that they think this stress is worthwhile since they then feel less stress during the chapter test. Table 11 shows two examples

from students about conflicting feelings. On the one hand they feel less stressed about the low-stakes FPG-tests but then more stressed at the chapter test. They also feel less confident about passing the chapter test when having low-stakes FPG-tests.

Original response	Translated response	Type of FPG-tests
<i>“Jag blir mer stressad över mina att hinna med alla frågor eftersom jag måste göra E-delen på provet. En annan nackdel är att jag inte är lika säker att jag kommer att klara proven med det nya sättet”</i>	<i>“I become more stressed over having time to do all questions because I have to do the E-part on the test. Another disadvantage is that I’m not as sure that I will pass the test with the new way”</i>	Low-stakes
<i>“Fördelar med e test som bara en avstämning är att man inte behöver se det som ett prov och då inte känner sig stressad av att man behöver plugga hemma till det. Nackdelar är att man istället behöver känna sig mer stressad till provet eftersom man inte får en andra chans på att visa förmågor på E-nivå.”</i>	<i>“Advantages with E test that only gives you a check up is that you don’t have to see it as a test and therefore don’t feel stressed that you need to study at home for it. Disadvantages is that you instead need to feel more stressed on the chapter test since you don’t get a second chance to show your abilities on E-level”</i>	Low-stakes

Table 11: Responses from students representing the views on negative feelings in connection to FPG-tests.

4.1.4 Study habits

The theme study habits includes students’ experience with FPG-tests having a positive effect on their study habits and motivation, as well as how they can affect them in a negative way. From both surveys, a total of 74 responses were identified as fitting into this category. Students express that they feel less motivated to study when they have low-stakes FPG-tests. Furthermore, students explained that they do not take low-stakes FPG-tests seriously since they don’t get rewarded for performing on them. Students also experience a decrease in motivation when they feel the pressure to study before the low-stakes FPG-tests. This is caused by students feeling obliged to study and prepare for low-stakes FPG-tests but don’t experience a feeling of being rewarded for their performance. Table 12 highlights two responses from students regarding feelings.

Original response	Translated response	Type of FPG-tests
<i>“Jag ser ingen fördel med det system vi har använt nu, det är slöseri med lektionstid som bättre</i>	<i>“I don’t see any advantages with the system we have used now, it is a waste of lesson time that could better be used</i>	Low-stakes

<i>läggs på genomgångar/eget arbete. Jag tycker inte jag får någon feedback från e testerna för hur det går för mig i kursen och det ger mig ingen motivation till att prestera.”</i>	<i>for walkthroughs/working on your own. I don't think I get any feedback from the FPG-tests on how I'm doing in the course and it gives me no motivation to perform”</i>	
<i>“Endast avstämning gör att jag inte känt lika hög press och motivation att lära mig. Chans till avklarad E-nivå hjälpte mig att inte prokrastinera och plugga direkt.”</i>	<i>“Low-stakes have made me not feel as much pressure and motivation to learn. The opportunity for passing E-level helped me not procrastinate and study right away”</i>	Low-stakes

Table 12 Responses highlighting motivation and study habits

The second response in Table 12 is also an example of how students feel high-stakes FPG-tests help them study more regularly. When they have the opportunity to be rewarded for their work their motivation increases, and they make sure they keep up with the teacher's lesson plans. Students also express that high-stakes FPG-tests help them feel motivated to learn the material during the course and not just for the test, see Table 13.

Original response	Translated response	Type of FPG-tests
<i>“Fördel är att man får motivation till att lära sig sakerna under kapitlen istället för att bara göra allt innan det stora provet. [...]”</i>	<i>“Pro is that you get motivated to learn the things during the chapters instead of just doing everything before the big test [...]”</i>	High-stakes
<i>“Jag tycker att e-test bidrar till att jag blir motiverad och att jag sprider ut mitt plugg över en längre period. [...]”</i>	<i>“I think that FPG-tests contribute to me being more motivated and that I spread out my learning over a longer period. [...]”</i>	High-stakes

Table 13: Responses from students regarding study habits.

4.2 High-stakes

A total of 139 students responded to the survey after being exposed to high-stakes FPG-tests. Figure 3 shows that 94,6% of students think they learn better from having high-stakes FPG-tests, while 5,4% think they learn more from having low-stakes FPG-tests.

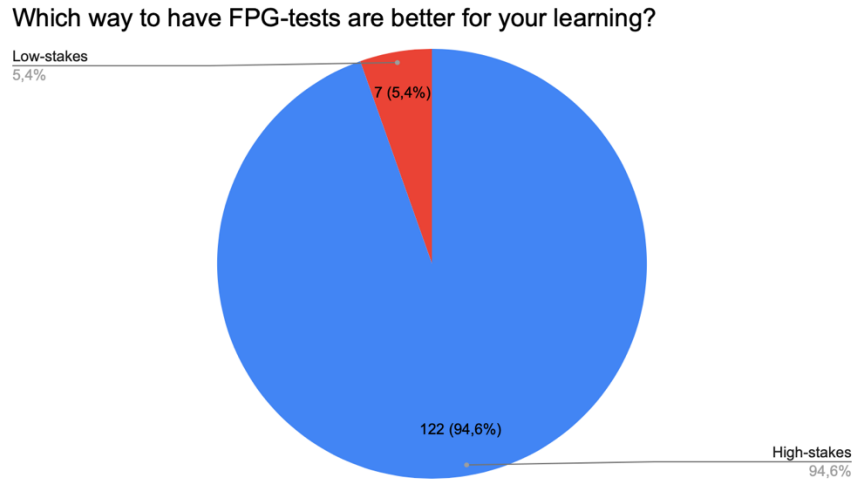


Figure 3: Which way students think they learn most from FPG-tests when having high-stakes FPG-tests.

In Table 14 percentages of students who passed the high-stakes FPG-tests can be found. Class Alfa and Gamma are the only two classes where more than half of the class passed the high-stakes FPG-tests. Classes Beta, Delta and Zeta are all below 50%. For class Epsilon an error occurred when collecting the data, a substitute teacher thought the FPG-tests were voluntary and only two students completed them which is why they only have 4,0% of students who passed.

Class	Percentages of students who passed high-stakes FPG-tests (%)
Alfa	86,2
Beta	40,0
Gamma	65,6
Delta	22,2
Epsilon	4,0
Zeta	28,6

Table 14: Percentages of students who passed high-stakes FPG-tests.

4.2.1 Follow-up work students feel they learn most from when having high-stakes FPG-tests

In Figure 4 the type of follow-up work students think they learn the most from when they have high-stakes FPG-tests can be found. Teacher-led assessment review is the type of follow-up work students feel they learn the most from, namely 47,8%. Self-assessment is closely followed at 44,2%, and peer-assessment is what 8,0% of students think they learn from.

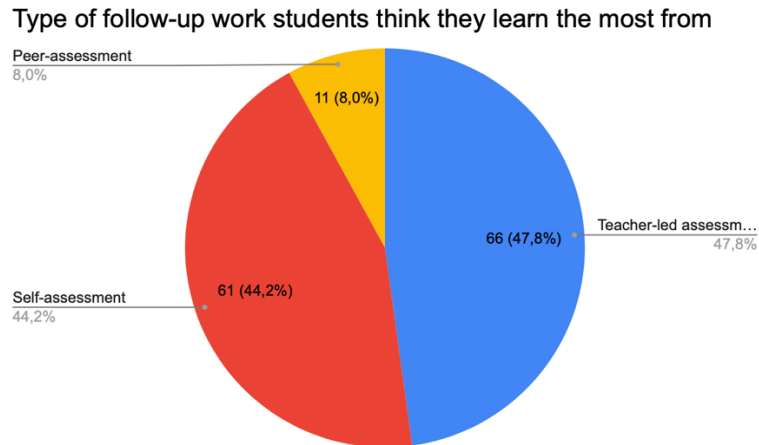


Figure 4. Type of follow-up work students think they learn the most from when having high-stakes FPG-tests.

This is illustrated in Table 15 where it is shown that students with lower grades prefer teacher-led assessment reviews, while students with higher grades, such as B's and A's prefer self-assessment. The group of students who prefer peer-assessment are the students who have the grade of D, which is the same result as when students had low-stakes FPG-tests.

Grade	Percentages Teacher-led assessment review (%)	Percentages self-assessment (%)	Percentages peer-assessment (%)
E	64,3	28,6	7,14
D	52,9	29,4	17,6
C	63,8	29,8	6,38
B	29,6	62,9	7,41
A	29,4	61,8	8,82

Table 15: Percentages of what kind of follow-up work students feel they learn the most from when having high-stakes FPG-tests sorted into grades.

To investigate whether or not the relationship between preferred follow-up work and grades is statistically significant when students have high-stakes FPG-tests, a chi-squared was conducted. This test revealed that $p < 0.05$, that there is a statistically significant relationship between preferred follow-up work and students' academic performance at the 5% level significance. The null hypothesis were therefore rejected and the alternative hypothesis accepted.

4.2.2 Optimal timing for feedback on high-stakes FPG-tests

Figure 5 is a circle diagram of when students think the optimal timing for feedback is for high-stakes FPG-tests. There is not a big difference between when students feel they learn the most from feedback. 51,8% of students think they learn more when feedback is being given immediately, while 48,2% think feedback should be delayed until the lesson after. This is almost the same response from students when they had low-stakes FPG-tests, with 47,1% respectively 52,9%.

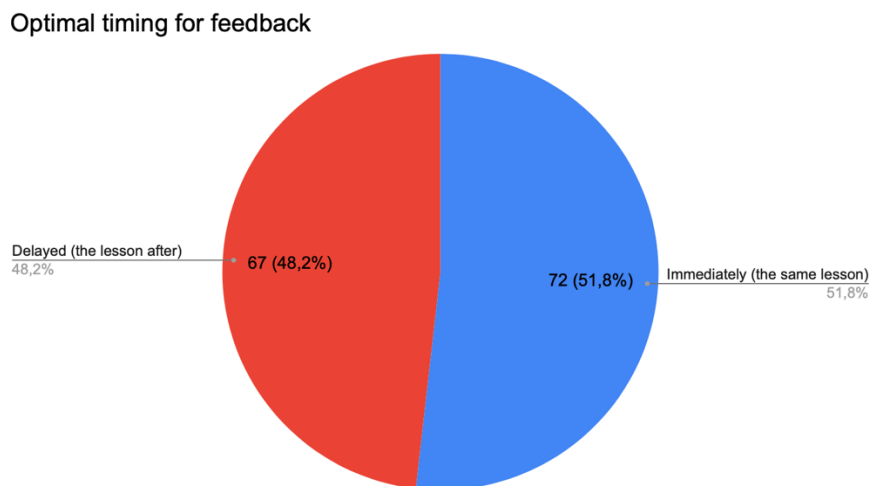


Figure 5: When students think feedback should be given after having high-stakes FPG-tests.

Table 16 summarises the information found in Figure 5 and shows them in percentages as well as sorted into grades. The groups of students with the grades E and D are both at 50% of when they think feedback should be given. The percentages of students who prefer feedback immediately increases as grades increase.

Grade	Percentages of students who prefer feedback immediately (%)	Percentages of students who prefer feedback to be delayed (%)
E	50,0	50,0
D	50,0	50,0
C	39,6	60,4
B	59,3	40,7
A	64,7	35,3

Table 16: Timing for feedback after high-stakes FPG-tests in percentages.

To investigate if a statistically significant relationship is present among students grades and when they think the optimal timing for feedback is, a chi-squared test was performed. The test

revealed that there is no statistically significant relationship between the optimal timing for feedback and students' grades. Consequently, the null hypothesis was accepted. Therefore, these results are probably due to chance.

4.3 Low-stakes

This chapter will present the survey responses from students when they had low-stakes FPG-tests. Figure 6 shows the responses from students when asked if they think high-stakes or low-stakes FPG-tests are better for their learning. Only 8,3% of students think they learn better from low-stakes FPG-tests while 91,7% think they learn more from having high-stakes FPG-tests.

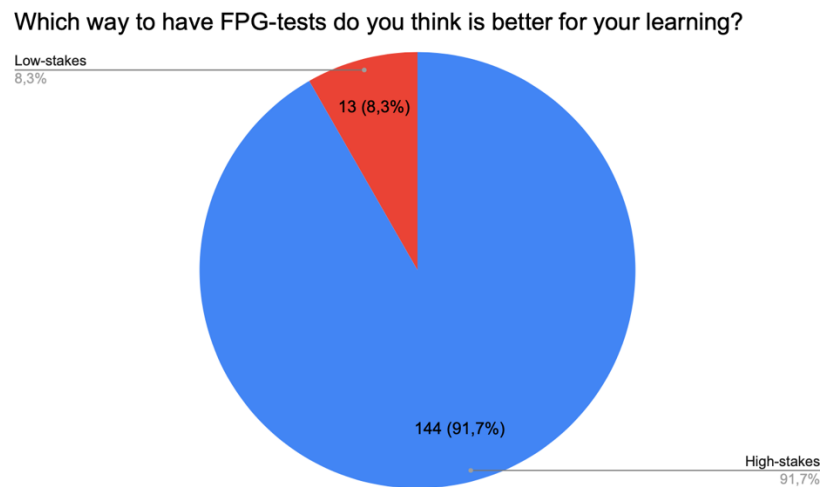


Figure 6: Which way students think they learn better from when they had low-stakes FPG-tests.

In Table 17 the percentages of students who passed the low-stakes FPG-tests can be found. All classes except two, Alfa and Gamma, had more than half of the class passing the FPG-tests when they were low-stakes. Class Delta is the class that had the highest pass frequency for students during low-stakes FPG-tests, while class Alfa was the class where fewest students passed the low-stakes FPG-tests.

Class	Percentages of students who passed low-stakes FPG-tests (%)
Alfa	41,4
Beta	53,3
Gamma	40,6
Delta	92,6
Epsilon	84,0
Zeta	64,3

Table 17: Percentages of students who passed low-stakes FPG-tests.

4.3.1 Follow-up work students feel they learn best from when having low-stakes FPG-tests

In Figure 7 the type of follow-up work students feel is best for their learning can be found. Teacher-led assessment review is the type that a majority of students feel they learn from, namely 57,3%, while 36,9% feel that self-assessment is the type of follow-up work they learn the most from. Only 5,7% of students think that peer-assessment is how they learn best when they have low-stakes FPG-tests.

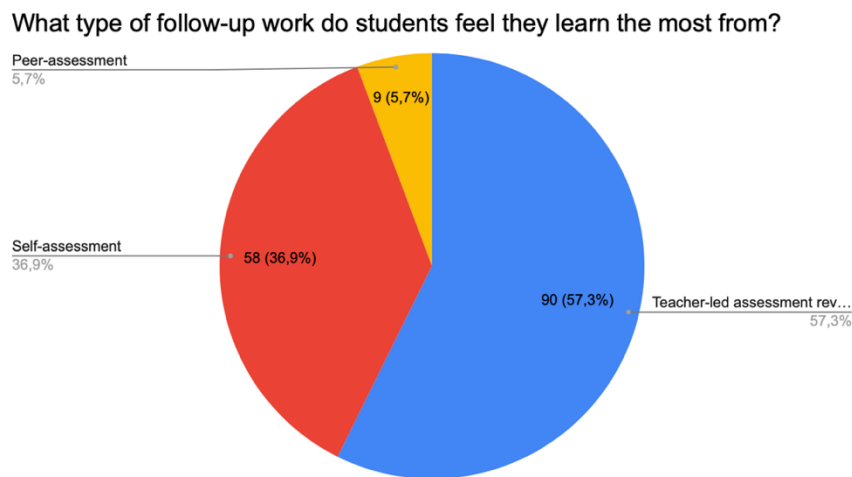


Figure 7: What type of follow-up work students feel they learn the most from when they have low-stakes FPG-tests.

In Table 18 percentages of students for each grade and which type of follow-up work they feel is best for their learning when having low-stakes FPG-tests can be found. Students with grades E and D prefer teacher-led assessment reviews, which then decreases with increasing grades, as the type of follow-up work students feel they learn best from. Students who have higher grades, i.e. B's and A's, mostly think they learn most from self-assessing. Peer-assessment is not as popular among students; students with grade D are the group where most students prefer peer-assessment, at 15,0%, followed by the group of students who have grade B at 6,67%.

Grade	Percentages Teacher-led assessment review (%)	Percentages self-assessment (%)	Percentages peer-assessment (%)
E	69,2	30,8	0
D	65,0	20,0	15,0
C	73,7	22,8	3,51

B	43,3	50,0	6,67
A	35,1	59,5	5,41

Table 18: Percentages of students who think they learn most from each follow-up work sorted into grades when having low-stakes FPG-tests.

To find out whether or not there is a statistically significant relationship between the type of follow-up work students think they learn most from and grades, another chi-squared test was conducted. The test revealed that $p < 0.01$, that there is a statistically significant relationship between preferred follow-up work and students' academic performance at the 1% significance level. Therefore, the null hypothesis that no statistical relationship could be found was rejected and the alternative hypothesis was accepted.

4.3.2 Optimal timing for feedback on low-stakes FPG-tests

Figure 8 show students' responses when asked for the optimal timing for feedback is for their learning when they have low-stakes FPG-tests. As can be seen, there is not a big difference in when students think the optimal timing for feedback is. Slightly more than half, namely 52,9% of students, think feedback should be given immediately, whereas 47,1% think it should be delayed until the next lesson.

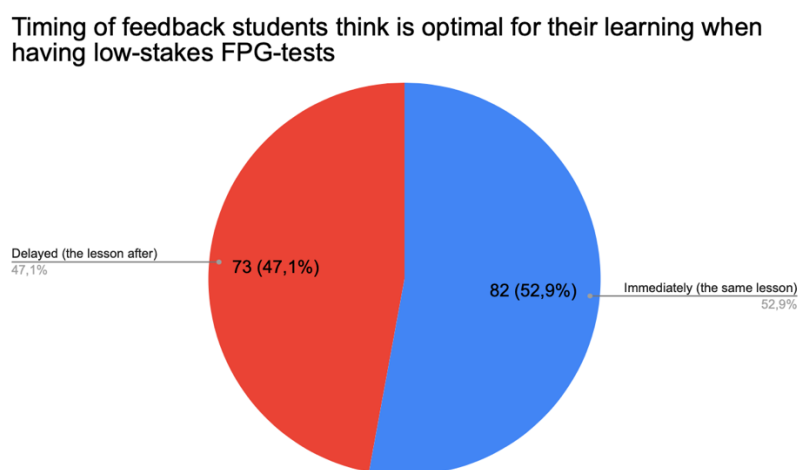


Figure 8: Timing of feedback students think is optimal for their learning when having low-stakes FPG-tests.

Table 19 illustrates that with increasing grades students also increasingly prefer to be given feedback immediately. In the groups of students with the grades E, D, and C, more than 50% prefer to be given feedback on the lesson after the low-stakes FPG-tests. This pattern was also found in the results from when students had high-stakes FPG-tests.

Grade	Percentages of students who prefer feedback immediately (%)	Percentages of students who prefer feedback to be delayed (%)
E	46,2	53,8
D	50,0	50,0
C	44,6	55,4
B	60,0	40,0
A	63,9	36,1

Table 19: Percentages of students' opinions on timing of feedback when having low-stakes FPG-tests sorted into grades.

To investigate if a statistically significant relationship is present among students grades and when they think the optimal timing for feedback is a chi-squared test was performed. The test revealed that $p > 0.05$, that there is no statistically significant relationship between the optimal timing for feedback and students' grades. The null hypothesis was consequently accepted. Therefore, these results are probably due to chance.

5. Discussion

From the results, it is evident that students prefer high-stakes FPG-tests over low-stakes FPG-tests. Students' perception is that they learn more from having high-stakes FPG-tests. The most prominent opinion among students is that they save time on the chapter test when they are given high-stakes FPG-tests. This results in students feeling calmer during the chapter test contributing to students believing they can perform better on the higher-grade questions. Students also expressed they feel more stressed before the high-stakes FPG-tests but that this stress is worthwhile because it helps them study more regularly and help them feel less pressure to perform on the chapter tests. Additionally, it is worth mentioning students' views on learning and performance. Numerous students experience a difficulty in separating their performance from their learning. Students express in both surveys that they learn more because they reached a higher grade. Furthermore, everyone who participated in the study has a grade of E or above which means that no one is failing the course. Teachers have also commented that they have seen an increase in grades since implementing FPG-tests compared to before. Since the students at the school are ambitious and generally get high grades, this could affect their view on learning, resulting in students associating their performance with learning.

Moreover, students express that low-stakes FPG-tests makes them feel less stressed compared to high-stakes FPG-tests. This could be due to low-stakes FPG-tests not being graded and therefore students don't experience performance anxiety before sitting through them. Students may feel that they don't have anything to lose when executing the low-stakes FPG-tests since they aren't being graded, resulting in them feeling less stressed. Students also emphasize that they remember the material better when they are exposed to low-stakes FPG-tests. This could be due to the fact they are being tested on the material twice, forcing them to study continuously. This is something that is presented in literature. Frequent testing helps students retain information better and therefore remember it for a longer period of time. When students instead partake in high-stakes FPG-tests there is a risk that they forget the fundamentals since they are only tested on the material once. There might also be a risk that students only prepare and study the essentials for the FPG-tests and then don't revise it at all in preparation for the chapter test. This means that even though they can pass the FPG-test they don't learn the material as deeply and therefore have problems with remembering and storing information long term. This was also presented as a risk in the literature, where students only memorise the material for the high-stakes FPG-tests and then don't repeat it, leading to them forgetting it faster. Thus, even though

students prefer high-stakes FPG-tests, they might actually learn more from having low-stakes FPG-tests as well as retain the information for longer and feeling less stressed.

Based on students' performance on the FPG-tests there seemed to be a higher passing rate on low-stakes FPG-tests compared to high-stakes FPG-tests. This could be the effect of students feeling less stressed when having low-stakes FPG-tests and were therefore able to absorb the material more efficiently during the lessons leading up to the chapter test. Another explanation for these results could be the area they were working with. Teachers at the school expressed that some students had low-stakes FPG-tests that were easier than when they had high-stakes FPG-tests since the chapter they were working with was easier. To find out if this is why students in greater extent passed the FPG-tests when they were low-stakes, this needs to be further studied and examined for a longer period of time, exposing students to low-stakes and high-stakes FPG-tests in more than one work area. It is also difficult to draw any conclusions on students' scores on their high-stakes FPG-tests. Class Epsilon had a substitute teacher who was under the impression that the FPG-tests were voluntary, which resulted in only two students actually doing the FPG-tests, leading to a misrepresentative result. Furthermore, since students' absence on the time of the FPG-tests wasn't taken into consideration, more students may have passed the FPG-tests than are highlighted in the results if they had not been sick. Perhaps, by ruling out students who missed one or more of the FPG-tests due to sickness or allowing them to take them from home, the results would be more reliable. Furthermore, students seem to perform well regardless of if the FPG-tests are high-stake or low-stakes. This could be due to the fact that students at the school are very ambitious. It would therefore be interesting to investigate if a more significant difference can be found between high-stakes and low-stakes FPG-tests among students who aren't as ambitious.

One explanation for obtaining these results could be that students are used to having high-stakes FPG-tests since that is what they did before this study started at the school. This could affect the results since students experience that something is being taken away from them when they are exposed to low-stakes FPG-tests. This change might also take some time to get used to for students which could make students not give it a proper chance and therefore be set on getting their high-stakes FPG-tests back, which could affect the results. Therefore, it would be interesting to evaluate low-stakes FPG-tests with new students who start at the school in August who aren't used to having high-stakes FPG-test. Then maybe, low-stakes FPG-tests would be taken more seriously among students and assist them in their learning process more effectively.

The result from the surveys also establish that students think either teacher-led assessment review or self-assessment is the type of follow-up work they learn best from in connection with FPG-tests, regardless of if they are high-stakes or low-stakes. In both cases, the results suggest a statistically significant relationship between the type of follow-up work students think they learn best from and their academic performance. Students with lower grades were more inclined to prefer teacher-led assessment reviews than students with higher grades who in a greater extent favoured self-assessment. Students who reach higher grades are typically more involved in their learning process and therefore can utilise self-assessing as a type of follow-up work in a more efficient way. High-achieving students also have a more developed self-efficacy, resulting in a deeper understanding of how they learn. Students who have higher grades may also have a higher confidence in their mathematical knowledge and therefore exhibit a higher level of self-trust. Self-assessment takes practice, and it forces students to become more active in their learning, which may be a behaviour high-achieving students already possess.

Students with lower grades preferred that the teacher goes through the solutions to the FPG-tests. One reason might be that these students don't feel as confident in checking through their mistakes on their own and experience a harder time critically thinking about what they could have done differently. Another reason for this, that is also presented in the literature, is that low-achieving students have a reduced metacognition and self-efficacy compared to high-achieving students. Something that is presented in literature is that teacher-led assessment reviews help students develop their metacognition and self-efficacy which might be an explanation to why students with lower grades experience they learn more from the teacher going through the answers than from self-assessment. Since high-achieving students already possess these skills, they might not benefit as much from teacher-led assessment reviews as students with lower grades and therefore prefer self-assessment.

Students expressed in greater regard that peer-assessment as follow-up work is better for their learning when having high-stakes FPG-tests compared to low-stakes FPG-tests. One student specifically said that they think it's easier to take in the feedback from their peers because they are on the same level. This is something that is presented in the literature as one of the benefits of peer-assessing. Students speak the same language and are able to transfer knowledge that the teacher might miss. One explanation for these results could be that students feel less motivated to discuss solutions with their peers when the FPG-tests are low-stakes since they already know

they have another chance to demonstrate their knowledge. Furthermore, students might not take the low-stakes FPG-tests, and the associated follow-up work as seriously since they aren't provided with confirmation from the teacher that they have reached the grade level E. Without this confirmation students might feel as they are not being rewarded for their performance further reducing their engagement and motivation to discuss the low-stakes FPG-tests with their peers. However, students take high-stakes FPG-tests more seriously since they have the opportunity to reach the grade level E which might increase their motivation to discussing solutions with their peers. Another aspect to consider regarding these results is the fact that students at the school are not as used to doing peer-assessment as they are to self-assessment and teacher-led assessment reviews. This is something the teachers at the school confirmed. Literature also highlights that for peer-assessment to be useful for students they need to practice. Therefore, students might need to be more exposed to having peer-assessment for them to experience the advantages it can have on their learning. To utilise the benefits of peer-assessment, teachers at the school should try exposing their students to this early and regularly to see if it has a positive effect on their learning. The teachers at the school can also keep varying the follow-up work to be able to cater to all their students at different grade levels.

There wasn't a big difference that could be identified in when students feel feedback should be given to optimise their learning when having low-stakes or high-stakes FPG-tests. What could be understood for both low-stakes and high-stakes FPG-tests are that students with higher grades preferred immediate feedback, whereas students with lower grades preferred feedback to be delayed. However, a statistically significant relationship could not be found between students grades and their preferred timing for feedback. One explanation for these results could be that high-achieving students also preferred self-assessment and therefore want to be given the opportunity to self-assess immediately after the FPG-tests. This, however, is the opposite of what was presented in the literature that said high-achieving students should be given delayed feedback. Although, even if students perceive that immediate feedback is more effective for their learning, they might still benefit from feedback being delayed. Perhaps if students got feedback a week after the FPG-tests instead it would have a larger impact on their learning. It could be that the next lesson was too short time to delay the feedback for students to show any significant results.

6. Conclusion

The aim of this thesis was to investigate different variations in FPG-tests with the purpose of finding out which way is most effective for increasing students' learning. Two types of FPG-tests were studied, and three variations of follow-up work, along with when the follow-up work should be practiced. Six classes from an upper secondary school in Gothenburg participated in the study that took place over the span of twelve weeks. The classes were divided into two groups and exposed to each type and all variations during the first six weeks and then they switched for the following six weeks, meaning all students tried all combinations. Students were given two surveys in total, one after each six-week period. Their responses were collected and analysed based on which type of FPG-tests students had alongside their scores on the FPG-tests.

Students' perception is that they learn better when they have high-stakes FPG-tests. The reasons they give are that they are able to focus on the harder questions. This leads to students feeling less stressed during the chapter test, but more stressed before the FPG-tests. High-stakes FPG-tests also lead to students feeling motivated to study and help reduce procrastination among students. Low-stakes FPG-tests on the other hand reduce stress connected to the FPG-tests among students and help them remember the material better, meaning they learn more. Although, students experience a decrease in study motivation for the FPG-tests and therefore might not see the benefits of having low-stakes FPG-tests, even though the results point to low-stake FPG-tests being better for their learning.

Based on students' scores, classes Alfa and Gamma are the only classes that perform better on high-stakes FPG-tests compared to low-stakes FPG-tests. All the other classes perform better on low-stakes FPG-tests. However, since there was an error in class Epsilon involving the substitute teacher this conclusion cannot be drawn from these results and needs to be further examined. Furthermore, it is important to keep in mind that if students missed one or more FPG-tests they were marked as failed, since they are not given an opportunity to redo the FPG-tests.

According to students' perceptions, students experience that they learn more when having either a teacher-led assessment review or through self-assessment. Students with lower grades prefer teacher-led assessment reviews, which could be due to them trusting teacher more than their

own abilities. Low-achieving students also have a reduced meta-cognition, which is increased through teacher-led assessment reviews, which in turn may explain why they feel they learn more from teacher-led assessment reviews. They also express the advantage of being shown step by step how to solve the problems. Students who have higher grades prefer self-assessment since it helps them reflect on their own mistakes and, in that way, helps them remember the material better. A minority of students express that peer-assessment is how they learn the most and this is mainly when they have high-stakes FPG-tests. Discussing solutions with peers and hearing different perspectives assist students in broadening their own perspectives. This relationship between type of follow-up work and students grade levels was shown to be statistically significant. The teachers can utilise this information and integrate all three types of follow-up work to be able to meet the needs of most of their students.

Student's opinions on when feedback should be given are almost the same when having high-stakes or low-stakes FPG-tests. There is not a big difference in when students feel feedback should be provided in order to learn the most. This is also consistent with what the literature says about the timing for feedback. One deviation from the literature, however, is that high-achieving students at the school prefer immediate feedback, whereas literature presents that high-achieving students should be given delayed feedback. The teachers at the school may use this information and vary when feedback should be given to cater to students at all grade levels.

Interesting future work that could be done in relation to this is to evaluate whether students with diagnoses benefit from having FPG-tests. One student with dyslexia expressed how high-stakes FPG-tests have been an aid for them, and this would be interesting to follow up on, as well as expanding to other diagnosis such as ADHD or autism. Since students at the school are ambitious and high-achieving it would be interesting to examine how FPG-tests affect students with lower grades or students who are at risk of failing mathematics. Finally, it would also be interesting to examine low-stakes FPG-tests on first year students when they haven't been exposed to high-stakes FPG-tests before.

References

- Barnes, M. (2017). Washback: Exploring what constitutes ‘good’ teaching practices. *Journal of English for Academic Purposes*. <https://doi.org/10.1016/j.jeap.2017.10.003>
- Bates, B. (2023). *Learning Theories Simplified* (3rd edition). Sage publishing.
- Biton, Y. (2025). Student Reflections on Peer Assessments: Benefits and Challenges in a Mathematics Class. *Educational Process International Journal*, 14(1). <https://doi.org/10.22521/edupij.2025.14.3>
- Bjork, R., & Ligon Bjork, E. (2025). *Research – Bjork Learning and Forgetting Lab*. Retrieved 23 January 2025, from <https://bjorklab.psych.ucla.edu/research/>
- Bjorkman, H. (2015, January 28). *Din 5-stegsguide till att analysera enkäter*. Netigate. <https://www.netigate.net/sv/articles/netigate-tips/din-5-stegsguide-till-att-analysera-undersoeningsresultat/>
- Bouchrika, I. (2025). *What is Mastery Learning Model? Definition, Principles and Examples in 2025*.
- Brosvic, G. M., & Epstein, M. L. (2007). *Enhancing learning in the introductory course*. Rider University.
- Cera, R., Mancini, M., & Antonietti, A. (2013). Relationships between Metacognition, Self-efficacy and Self-regulation in Learning. *ECPS - Educational, Cultural and Psychological Studies*, 7, 115–141. <https://doi.org/10.7358/ecps-2013-007-cera>
- Demir, C. G., & Keleş, Ö. K. (2021). The impact of high-stakes testing on the teaching and learning processes of mathematics. *Journal of Pedagogical Research*, 5(2), 119–137. <https://doi.org/10.33902/JPR.2021269677>
- East Carolina University. (2024). *How Does Human Memory Work?* <https://onlineprograms.ecu.edu/blog/how-does-human-memory-work/>
- Emeka, C., Herman, G., & Zilles, C. (2023). Leveraging Second-Chance Testing to Improve

- Students' Outcomes. *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 2*, 50–51. <https://doi.org/10.1145/3568812.3603490>
- Esaiasson, P., Gilljam, M., Oscarsson, H., Towns, A., & Wägnerud, L. (2017). *Metodpraktikan—Konsten att studera samhälle, individ och marknad* (Femte upplagan). Wolters Kluwers.
- Favero, T. G., & Hendricks, N. (2016). *Student exam analysis (debriefing) promotes positive changes in exam preparation and learning*. <https://doi.org/10.1152/advan.00060.2016>
- Hattie, J., & Timperley, H. (2007). *The Power of Feedback*. <https://doi.org/10.3102/003465430298487>
- Janison Assessments. (2022). *High stakes vs low stakes testing—Key differences to know*. <https://www.janison.com/resources/post/high-stakes-vs-low-stakes-testing/>
- Kangaslampi, R., Asikainen, H., & Virtanen, V. (2022). Students' perceptions of self-assessment and their approaches to learning in university mathematics. *LUMAT: International Journal on Math, Science and Technology Education*, 10(1), Article 1. <https://doi.org/10.31129/LUMAT.10.1.1604>
- Kemp, K. (2021). *Test Corrections Appear To Benefit Lower-Achieving Students in an Introduction to Biology Major Course: Results of a Single-Site, One-Semester Study*. https://journals.asm.org/doi/epdf/10.1128/jmbe.00122-21?src=getftr&utm_source=scopus&getft_integrator=scopus
- Morphew, J. W., Silva, M., Herman, G., & West, M. (2020). Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering. *Applied Cognitive Psychology*, 34(1), 168–181. <https://doi.org/10.1002/acp.3605>
- Mullet, H. G., Butler, A. C., Verding, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*.

<https://doi.org/10.1016/j.jarmac.2014.05.001>

Muzsnay, A., Zábó, C., Szeibert, J., Bernáth, L., Szilágyi, B., & Szabó, C. (2024). How do testing and test-potentiated learning versus worked example method affect medium- and long-term knowledge in abstract algebra for pre-service mathematics teachers? *European Journal of Psychology of Education*, 39(4), 4395–4415. Scopus. <https://doi.org/10.1007/s10212-024-00869-8>

Panadero, E., Jonsson, A., & Botella, J. (2017). *Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses*. <https://doi.org/10.1016/j.edurev.2017.08.004>

Penk, C., Pöhlmann, C., & Roppelt, A. (2014). The role of test-taking motivation for students' performance in low-stakes assessments: An investigation of school-track-specific differences. *Large-Scale Assessments in Education*, 2(1), 5. <https://doi.org/10.1186/s40536-014-0005-4>

Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Chapter One—Ten Benefits of Testing and Their Applications to Educational Practice. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 55, pp. 1–36). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00001-6>

Schüttpelz-Brauns, K., Hecht, M., Hardt, K., Karay, Y., Zupanic, M., & Kämmer, J. E. (2020). Institutional strategies related to test-taking behavior in low stakes assessment. *Advances in Health Sciences Education*, 25(2), 321–335. <https://doi.org/10.1007/s10459-019-09928-y>

Simzar, R. M., Martinez, M., Rutherford, T., Domina, T., & Conley, A. M. (2015). Raising the stakes: How students' motivation for mathematics associates with high- and low-stakes test achievement. *Learning and Individual Differences*, 39, 49–63. <https://doi.org/10.1016/j.lindif.2015.03.002>

Smith, H., Ngo, V., Sales, A., Closser, A. H., Chan, J., & Ottmar, E. (2022). To Wait or Not to Wait: Adding to the Debate on Immediate Versus Delayed Feedback. *International Society of The Learning Sciences*.

- Soderstrom, N. C., & Bjork, E. L. (2023). Pretesting Enhances Learning in the Classroom. *Educational Psychology Review*, 35(3). Scopus. <https://doi.org/10.1007/s10648-023-09805-6>
- Svensson, Å., & Vesterberg, F. (2024). *Frequent Passing Grade-tests—A study on the impacts of frequent tests at passing grade-level on students and teachers in upper secondary school mathematics*. Chalmers University of Technology.
- University of Texas at Austin. (2025). *Practice and Frequent Testing*. <https://ctl.utexas.edu/frequent-testing>
- Walck-Shannon, E. M., Cahill, M. J., McDaniel, M. A., & Frey, R. F. (2019). Participation in Voluntary Re-quizzing Is Predictive of Increased Performance on Cumulative Assessments in Introductory Biology. *CBE—Life Sciences Education*, 18(2), ar15. <https://doi.org/10.1187/cbe.18-08-0163>
- Zaveri, A. (2024). Chi-två-test: Förstå och tillämpa detta statistiska verktyg - Mind the Graph Blog. *Mind the Graph Blogg*. https://mindthegraph.com/blog/sv_se/chi-square-test/

Appendix

Appendix 1: Survey given to students

Enkät E-tester

2025-05-07 10:38

Enkät E-tester

Vi kommer i år att undersöka E-tester och kolla vilka aspekter som är viktiga för att de ska vara framgångsrika. Vi vill därför veta vilka erfarenheter ni har från E-tester.

Enkäten är pseudoanonym, ditt namn står inte direkt i svaret men kan tas fram via en kod.

*** Anger obligatorisk fråga**

1. Vad är din kod *

2. Samtycker du till att dina svar används till en studie på Chalmers? *

Detta behövs för att pseudoanonyma data ska kunna användas i ett examensarbete på Chalmers.

Om du vill veta mer finns det information här:

<https://docs.google.com/document/d/104VC9lRgDzkEiHsKTlfrv9sU0vJF3kpDNdJZR1h0FfE/edit?usp=sharing>

Markera endast en oval.

Ja

Nej

Två sätt att ha E-test

E-test kan antingen vara att du har möjlighet att bli klar med E-nivå eller att det bara är en avstämning då du testar dina kunskaper. Det senaste avsnitten har vi jobbat med ett av dessa sätten.

6. Kommentera gärna ditt svar:

Efterarbete till E-test

7. Vilket av dessa alternativ av efterarbete tror du är bäst för ditt lärande? *

Markera endast en oval.

- Självbedömning
 Genomgång av läraren
 Kamratbedömning

8. När tror du det är bäst för lärandet att ha efterarbetet? *

Markera endast en oval.

- Samma lektion
 Lektionen efter



CHALMERS
UNIVERSITY OF TECHNOLOGY