



GÖTEBORGS UNIVERSITET
HANDELSHÖGSKOLAN

Does a changed grading scale affect mean GPA?

A descriptive analysis of the 2011 grading scale reform in the
Swedish upper secondary school

Bachelor thesis in economics (15 credits)

Spring 2016

Department of Economics

Authors

Elin Bengtsson

Victoria Andersson

Supervisor

Andreea Mitrut

Abstract

That students receive accurate grades is of high importance, both for employers that want to hire productive workers and for university admission. Grades can also be used as a motivation for students to perform better, and different formulations of the grading scale is likely to have different motivational effects. Still, very little research has been done on the effects of the changed grading scale in the Swedish upper secondary school of 2011. The aim of this thesis is to examine changes in mean grade point average (GPA) in the Swedish upper secondary school after the implementation of the new grading scale, and to examine if there are any indications of the new grading scale having any motivational effects. We find that the mean GPA has decreased after the implementation of the new grading scale. We also find indications on that higher achieving programmes have had a smaller decrease in mean GPA, which could indicate that the new stricter grading scale has had a motivational effect on high achieving students.

Keywords: grading practices, effects on GPA, educational incentives, student performance, Swedish upper secondary school

Acknowledgements

We are very grateful for all the advice, help and feedback received from our supervisor Andreea Mitrut during the writing of this thesis.

Contents

Abstract	1
List of Tables.....	3
1. Introduction	4
2. Institutional background.....	5
2.1 The old grading scale	5
2.2 An upper secondary school reform	5
3. Literature review	7
3.1 Grades as a signal for human capital.....	7
3.1 Effects of a remodelled grading scale	7
3.4 Earlier studies on the new Swedish grading scale.....	8
4. Data and identification strategy.....	9
4.1 Data	9
4.2 Identification strategy.....	12
5. Results and analysis	13
5.1 Overall regressions	13
5.2 Programmes with an above average GPA.....	15
5.3 Programmes with a below average GPA.....	17
5.4 Programmes with higher GPA	18
5.5 Control for time trend.....	20
6. Discussion	21
6.1 Aggregated data.....	21
6.2 Missing data in the database SiRiS	21
6.3 Omitted variable bias	21
6.4 The general school reform.....	22
References	24

List of Tables

4.1 Summary statistics

5.1 Overall regressions

5.2 Regression for programmes with higher than average GPA

5.3 Regression for programmes with lower than average GPA

5.4 Regression for programmes with GPA above the 75th, 90th and 95th percentiles

5.5 Control for time trend

1. Introduction

Research show that students' grades can be a good indicator of their abilities and that there is a significant predictive ability of high school grades on long-term productivity (Miller 1998). This indicates that students' grades can be a useful signal for employers that want to hire productive workers. For this reason it is important that this indicator is reliable. If grades are able to accurately show the abilities of each student, then grades can be a beneficial benchmark for future employers that want to hire new employees, or for university admission that are looking for candidates with suitable prior knowledge. Research also suggests that a broader grading scale can have an impact on students' motivation (Becker & Rosen 1992, Bressette 2002) and that a stricter grading can boost students' to work harder (Betts & Grogger 2003, Johnson & Beck 1988).

Although the reliability of the students' grades is of high importance, there has been little quantitative research about the effects of the changed grading scale in the Swedish upper secondary school in 2011. This thesis is to our knowledge the first attempt to examine any change in students' results after the implementation of the new grading scale, which includes additional grade levels and aims to be stricter in comparison to the old one. The thesis aims to examine changes in mean grade point average (GPA) in the Swedish upper secondary school after the grading scale change, and to examine whether there are any indications of the new grading scale having an impact on students' motivation.

To test whether the new grading scale has had any impact on students' results, we compare mean GPA before and after the new grading scale was implemented. The mean GPA is recorded on an aggregated level, so no individual results can be discerned. Instead of each student's mean GPA, our dataset provides the mean GPA of different programmes in different schools¹. When controlling for the share of students with highly educated parents, the share of women and the share of students with an immigrant background in every programme we find that the mean programme GPA has decreased after the new grading scale was implemented. We also find indications that, among the 50 percent highest achieving programmes, the higher achieving the programme is, the smaller is their decrease in mean GPA. This could indicate that the new stricter grading scale motivates already high achieving students, which is in line with previous research.

This paper continues with a short explanation of the policy change in section 2, followed by a literature review in section 3. Section 4 presents an account of the data used in the regressions, and

¹ As an example, one observation is the mean GPA of the Natural Sciences Programme of the Hvidfeldska gymnasium in Gothenburg.

section 5 shows the results from the regressions and a discussion of these results. Section 6 provides a discussion of possible shortcomings in the thesis and section 7 provides a conclusion.

2. Institutional background

2.1 The old grading scale

The previous grading scale was in use between 1994 and 2011 and used four different grades, IG (fail), G (pass), VG (pass with distinction) and MVG (pass with special distinction). The students were graded based on each course's national proficiency requirements that identified the knowledge the student needed to be able to show to earn the different grades. The grades were to be used to signify to what extent each student had reached each course's knowledge requirements. This was meant to be used as a signal to the student's themselves, their parents, prospective employers and higher educational institutions, where they could be used to rank students who applied to programmes with more applicants than places. Despite the grades being used as a ranking system like this, the government emphasized that this should not be more important than the grades' other functions, like showing to what extent the students' had reached the knowledge requirements. More grades than four were not deemed necessary. (SOU 1992/93:250)

When applying for higher education after upper secondary school, each student's grades were recalculated using a formula to get their grade point average. When calculating the GPA, the grades were given the values 0 for IG, 10 for G, 15 for VG and 20 for MVG. (SOU 1992/93:250)

2.2 An upper secondary school reform

In 2011 the Swedish government implemented the reformed grading scale for the upper secondary school. The old grading scale resulted in difficulties for the teachers to grade students that lay between two grades. This led to teachers sometimes grading students with a higher, or lower, grade than what they actually had earned, and one aim of the new grading scale was to deal with this problem. Another main motive for the new grading scale was to boost the students' motivation. A more differentiated grading scale would make it easier for students to distinguish themselves from each other, which would make them motivated to study harder. The new grading scale also aimed to be stricter than the old one and make it more difficult to earn higher grades, which was also meant to motivate students. (Ds 2008:13)

The reformed grading scale has six different grades, A, B, C, D, E and F, where A-E represents a passed result with A being the highest grade, and F representing a failed result. The purpose of the grades is to reflect the quality of the student's knowledge with regard to the proficiency requirements for the course, just as with the old grading scale. (Ds 2008:13)

For the grades E, C and A there are a number of specific proficiency requirements for each course that the student has to fully pass to earn the grade in question. To earn the grades B or D, the student must fully live up to all the proficiency requirements for the grade below, and predominantly live up to the grade above. This means that to earn the grade B the student has to fully live up to the proficiency requirements for the grade C and predominantly live up to the requirements for the grade A. If the student does not reach all knowledge requirements for grade E, the student will be graded with an F. When calculating the final grade, the course grades are given following values: 0 for F, 10 for E, 12,5 for D, 15 for C, 17,5 for B, 20 for A. (Ds 2008:13)

Hence, the new and old grading scale does not correspond. In the new scale the criteria for A, C and E must be fully met, which was not required to reach VG or MVG in the old system. Students that would have earned the grade MVG in the old system could with the new system earn the grade A or B. This means that it is more difficult to earn a higher grade with the new grading scale, which is in line with the government's intention. The government states that the new grading scale encourages students to put more effort into their studies and that it will increase the students' motivation. (Ds 2008:13)

The first cohort to receive their final grade with the new grading scale graduated in 2014. Cohorts that begin their education before 2011, and hence received their final grade before 2014, were graded with the old grading scale. (National Agency for Education 2016)

In addition to the grading scale change, the upper secondary school reform of 2011 also brought further changes to the school system. One of these changes was the formulation of the national programmes. Some of the programmes were altered, some were merged with other programmes, some new were added and some were removed. (National Agency for Education 2016). This is discussed further in section 4.

3. Literature review

3.1 Grades as a signal for human capital

Human capital theory states that employers are looking for employees with a high level of productivity and ability (Ehrenberg & Smith 1994, Miller 1998) and that employers will pay employees according to their productivity (Borjas 2016). However, it might be difficult for an employer to know a newly hired employee's true productivity, since the employer does not have all the information about the employee. For employers this could be especially difficult regarding younger workers with less working experience and few employer references. (Borjas 2016)

A way for employers to deal with this lack of information about the workers' productivity could be to look high school grades, since research show that grades can be a good indicator of students' productivity (Miller 1998). Miller has found that high school grade has a significant and strong effect on long-term productivity, measured as earnings nine years after graduation. By modelling grades or another achievement as an indicator of productivity, employers would be able to make more informed hiring decisions. (Spence 1973)

Hence, high school grades can be used as a signal to employers that are looking for good workers. The theory of signalling means that a signal can be used to reliably convey information between two parties who have asymmetric information (Borjas 2016). Applied in a labour market situation, this could be between an employer and a prospective employee, since employers usually do not have a trustworthy indicator of employees' productivity. A signal would therefore decrease this information asymmetry (Borjas 2016).

Applying the signalling theory to the grading scale, one could argue with the same assumptions as above that more levels on a scale could act as a motivation for students to achieve a higher grade and receive a better signal. This would also be in line with the government's motivation for a more differenced grading scale. (Ds 2008:13).

3.1 Effects of a remodelled grading scale

Previous literature states that added grade levels can work as an incentive to work harder, since the additional levels can differentiate students that previously would have reached the same grade as each other. Becker and Rosen (1992) argue that multiple grade levels will get a behavioural response from the students, since they will want to reach the higher available grade and distinguish themselves from others. Bressette (2002) has also found that a grading scale with additional levels

leads to better differentiation of students, as well as a motivational effect. When a grading scale has fewer levels the risk of grade compression increases, which means that it becomes more difficult to distinguish student with different abilities from each other. A proposed solution to grade compression is a larger grading scale (Bressette 2002) which teachers can use to differentiate students with different abilities. Research has also found an increase in grade reliability as the number of levels on a grading scale increases (Singleton & Smith 1978). However, one study of the implementation of a more differentiated grading scale found a negative effect on student GPA, as well as a negative perception of the scale among students (Edgar 2014).

In addition to having added grade levels, the new grading scale is also stricter than the old one. Literature argues that a strict grading scale can have both positive and negative effects on school achievements, depending on the individual student. Betts and Grogger (2003) come to the conclusion that students become more motivated by stricter grading standards, although it has a bigger effect on students who already have high grades. Johnson and Beck (1988), have similarly found that a stricter grading scale led to higher test scores. However, they have also found that this is only true to a certain level. When the standards are too high, the benefit of reaching a higher grade will be outweighed by the cost of additional study time. The marginal cost of studying can be too high and the marginal benefit of the higher grade can be too low to be motivating (Elikai & Schuhmann 2010). A strict grading scale can also lead to lower performing students giving up, since they perceive that the higher grade is unattainable (Betts 1997). In conclusion, there is a strong support among existing literature for a positive relationship between stricter grading standards and students' performance. The effect on lower achieving students, on the other hand, is not as clear.

3.4 Earlier studies on the new Swedish grading scale

As mentioned in the introduction, quantitative research regarding the effects of the 2011 grading scale change has been very limited. There have been some qualitative studies from the National Agency of Education (Skolverket) that address students' and teachers' attitudes on the new grading scale. These studies find that there is a "threshold-effect" where you are unable to earn the higher grades since, for example, you have to get an A on all knowledge criteria for a course in order to get an A as a final grade (National Agency for Education 2016). Therefore, it is now more difficult to earn the highest grade than it was before, which is in line with the motives of the government (Ds 2008:13). This can be considered to work in favour of the signalling theory effect, since the students who now receive an A are even more likely to be high ability students than the students

who received MVG with the prior scale. The same goes for the grade C in comparison to VG. However, this could also offset the motivating incentive of the higher grades for some students, since the marginal cost of reaching a higher grade might outweigh the marginal benefit.

4. Data and identification strategy

4.1 Data

Our data is collected from the database SiRiS, which is provided by the National Agency for Education. The data we are using provides information on the students in the third grade of the Swedish upper secondary school, and is ranging from years 2010 to 2014. Since the first cohort to graduate with the new grading scale received their final grade in 2014 our dataset contains four cohorts that were graded with the old grading scale, the cohorts of 2010-2013, and one cohort that was graded with the new grading scale, the cohort of 2014.

The data is aggregated and provided on programme level. The Swedish upper secondary school now has 18 different national programmes that can be either university preparatory or vocational. Prior to the upper secondary school reform that was explained in section 2.2, the upper secondary school had 17 national programmes that were formulated slightly differently from how the new programmes are formulated. Since some of the programmes have been changed, some no longer exist and some have been added, we had to consider this in our comparison of the programme results before and after the grading scale reform. Programmes that were merged with other programmes or significantly altered were not included in the dataset since our perception is that they are no longer comparable. The reason for this is that we do not know how the students would have chosen if the new programmes would have existed before the policy change, and we believe that the observations might would have looked different. Since we do not know how this would have biased the estimates, we excluded all programmes that we believe are not comparable before and after the policy change. One exception to this is the inclusion of the Electricity and Energy Programme. This programme was prior to the reform two different programmes: the Electricity Programme and the Energy Programme. Our perception is that the students that attended these two programmes are the same type of students that now attend the Electricity and Energy Programme, which is why we have merged the result for these two programmes to compare with the result of the Electricity and Energy Programme.

Due to these changes to the programmes we have chosen to include ten out of the 18 national programmes in our data set. These are the programmes that look the most similar before and after the reform. We have also compared the numbers of students in each programme before and after the school reform, to get an indication on whether the programmes we have chosen to compare resembles each other before and after the reform regarding what type of student they attract. The university preparatory programmes in our dataset consist of the Arts Programme (ES), the Natural Science Programme (NA), the Social Science Programme (SA) and the Technology Programme (TE). The vocational programmes in our dataset are the Building and Construction Programme (BA), the Child and Recreation Programme (BF), the Electricity and Energy Programme (EE), the Handicraft Programme (HV), the Health and Social Care Programme (VO), the Hotel and Tourism Programme (HT), the Natural Resource Use Programme (NB) and the Vehicle and Transport Programme (FT). Prior to the school reform some of these programmes had slightly different names.

Our dataset includes 5,398 individual programme observations distributed across 535 schools. Some programme observations available on SiRiS have been excluded from our dataset as they did not provide information about GPA for all the years, for example observations from newly established schools. In the regressions we control for the share of students with highly educated parents, the share of women and the share of students with immigrant background in every programme.

Below follows a description of the dependent variable GPA and the independent variables:

GPA

GPA is an abbreviation of grade point average, and is the average GPA for each programme. This is the final grade received the last year of upper secondary school. This is calculated, as mentioned above, by converting each student's course grades to points using converting tables, and then calculating the mean GPA for each programme. The unit of this variable is measured in grade points.

After

After is a dummy variable that takes the value 1 for the cohort that received their final grade during 2014, and 0 for the cohorts that received their final grade during the years 2010-2013.

Share of students with highly educated parents

The control variable *share of students with highly educated parents* is the share of students with highly educated parents in a programme. Having highly educated parents, as defined by the database SiRiS, means having at least one parent with a higher than upper secondary school education. According to Barnombudsmannen (2016), the education of a student's parents has a big impact on the student's educational achievements. This means that a programme with a higher share of students with highly educated parents might make the estimates positively biased when not controlled for. In the regressions the variable is called "educated_parents".

Share of women

The control variable *share of women* is the share of women in each programme. According to the Ministry of Education and Research (2004), women in upper secondary school have a higher GPA than men across all programmes. This means that a programme with a higher share of women might make the estimates positively biased when not controlled for. In the regressions the variable is called "women".

Share of students with immigrant background

The control variable *share of students with immigrant background* is the share of students with an immigrant background in a programme. The definition of immigrant background as used by the database SiRiS is being born outside of Sweden or having both parents born outside of Sweden. According to Barnombudsmannen (2016), having an immigrant background has a slightly negative impact on education. This means that a programme with a higher share of students with immigrant backgrounds might make the estimates negatively biased when not controlled for. In the regressions the variable is called "immigrant_background".

Table 4.1 Summary statistics

Variable	Before					After				
	Obs	Mean	Std dev	Max	Min	Obs	Mean	Std dev	Max	Min
<i>GPA</i>	4,248	14.37	.03	9.4	18.4	1,150	14.03***	.04	12	100
<i>Educated_parents</i>	3,924	52.35	18.06	12	100	1,094	51.70	18.00	16	96
<i>Women</i>	3,611	59.49	22.08	5	100	993	57.80**	23.38	5	100
<i>Immigrant_background</i>	1,943	23.14	.38	3	100	680	25.65***	.66	2	96
*** p<.01, ** p<.05, *p<0.1										

We performed a t-test in order to find whether the mean GPA and the controls are statistically significant before and after the implementation of the new grading scale, which would give us an indication on whether the new policy has had any impact on mean GPA. The null-hypothesis for the t-test for mean GPA was that the policy had led to zero change. Since the p-value was .0000 we can reject this null-hypothesis. In this case, it seems that the new grading policy is correlated with a decrease in mean GPA. However, these tests only show that there is correlation between the changing of the grading policy and the differences in mean GPA, they do not provide any causality. Table 4.1 shows that there is no significant change in the share of students with highly educated parents. The share of women has had a significant decrease while the share of students with an immigrant background has had a significant increase.

When only looking at GPA, there are 5,398 observations ranging from year 2010 to 2014. When controlling for the share of students with highly educated parents there are 4,835 observations. When controlling for the share of women there are 4,425 observations and when we control for the share of immigrants there are 2,552 observations for the same years. The reason for this is that the database SiRiS does not provide data when the result is based on fewer than 10 students. This missing information could create some sample bias, which is discussed in section 6.2.

4.2 Identification strategy

To test whether the implementation of a new grading scale has had any effect on mean GPA a before-after estimation will be used. Using a before-after type estimation means having to make a strong assumption – that in the absence of the policy change, nothing would have happened to the dependent variable before and after the time of the policy change. This means that other things that could have affected the grades would have been the same before and after the policy had been implemented. Although strong, the assumption could be plausible in this case. However, some factors that could affect GPA, such as students' effort and motivation might differ somewhat from year to year, some being more or less motivated than others. In addition, teacher behaviour might not be exactly the same from year to year and class sizes might also differ slightly. These differences might have impacted the results and could cause a bias in our estimates. However, macroeconomic conditions during the period have been fairly stable, and other regional economic characteristics are likely to have been similar as well, which means that the schools' economic conditions are likely to be similar before and after the policy change. Hence, it is probable that the schools' conditions and other factors that might affect programme GPA are fairly similar before and after the policy change and that a before-after analysis is applicable.

5. Results and analysis

In order to try to find the effects of the new grading policy an econometric regression analysis was done. Since panel data was used there was a choice between using a fixed effects regression model or a random effects regression model. Using a Hausman test² indicated that a fixed effects regression model would be better in this case, since in our data the programmes have unique attributes that do not vary over time and are not random variation. A fixed effects model only uses within-group variances when making the estimates; therefore, time-constant heterogeneity is controlled for. This could prevent omitted variable bias, which could be large in this case since data on control variables that could be of importance are missing. In our fixed effects regression, one “group” indicates one programme, for example the Natural Science Programme at one school. When doing the regressions, clustering of standard errors on school level was used in order to control for heteroscedasticity in the standard errors and make the standard errors robust to within-school correlation. Performing a Wooldridge test³ for autocorrelation in the data also showed that there was no first-order autocorrelation.

In this chapter we will provide the econometric result using regression tables. The regressions displayed are a selection of generated regressions that contribute to the understanding of the results.

5.1 Overall regressions

The linear regression including controls, is

$$GPA_{it} = \beta_0 + \beta_1 after_{it} + \beta_2 educatedparents_{it} + \beta_3 women_{it} + \beta_4 immigrantbackground_{it} + \varepsilon_{it}$$

where GPA is the dependent variable, β_0 is the constant, β_1 , β_2 , β_3 and β_4 represent the coefficients for the variables *after*, *educated_parents*, *women* and *immigrant_background*. i is the individual programme, t is the year and ε is the error term that captures all other unobservable factors that might have an impact on the dependent variable GPA . To find the effect of the new grading scale on GPA we start the analysis with an overall linear regression without controls that includes all the 5,398 programmes in our dataset.

² The null-hypothesis “difference in coefficients is not systematic” can be rejected with a p-value of 0.000

³ The null-hypothesis “there is no first-order autocorrelation” cannot be rejected with a p-value of 0.127

Table 5.1 Overall regressions

Variables	<i>1. GPA</i>	<i>2. GPA</i>	<i>3. GPA</i>	<i>4. GPA</i>	<i>5. GPA</i>	<i>6. GPA</i>
<i>After</i>	-.2980***	-.3482***	-.4343***	-.3869***	-.4688***	-.5542***
	(.0315)	(.0318)	(.0321)	(.0399)	(.0403)	(-.5542)
<i>Educated_</i> <i>parents</i>		.0128***			.0123***	
		(.0021)			(.0035)	
<i>Women</i>			.00448***		.0010***	
			(.0016)		(.0025)	
<i>Immigrant_</i> <i>background</i>				-.0266***	-.0238***	
				(.0031)	(.0032)	
Constant	14.3590** *	13.8241** *	14.3927** *	15.2926***	14.1164** *	14.8538** *
	(.0067)	(.1138)	(.0917)	(.0710)	(.2936)	(.0109)
Observations	5,398	4,835	4,425	2,552	2,394	2,394
R2	0.0350	0.0634	0.0844	0.1805	0.2205	0.1521
Standard error in parenthesis, clustered on schools. *** p<.01, ** p<.05, *p<0.1						

In table 5.1, regression one shows a significant decrease in mean GPA with .30 grade points after the implementation of the new grading scale. This means that the average GPA for all students has decreased by about 0.30 grade points after the changing of the grading policy. However, this regression is done without any controls and is likely affected by omitted variable bias.

Following regression one are regressions two to four that show the change in GPA when controlling for the share of students with highly educated parents, the share of women and the share of students with an immigrant background, respectively. The fifth column shows the change in GPA when including all of the control variables, and the coefficient for the new grading scale policy is now about -.47 grade points. This means that, when controlling for the different student characteristics, the mean GPA has decreased by about 0.47 grade points. This indicates a decreasing effect of the policy on programmes' mean GPA. The table also shows that all of the controls are significant at a 1 % level.

Since we have fewer observations when including all the control variables than when not using controls, we added one regression that uses the same observation sample as the regression that includes all of the controls, regression five, but without including the controls in the new regression. This regression is shown in column six. This regression is included to find out if the change in the coefficient for the variable "after" was different in the different regressions because they used different control variables, or because it was a different sample. Regression six shows a bigger decrease in mean GPA than regression one, which includes all observations. This indicates that it is the sample, rather than the controls, that affect the change in GPA in column five. The sample with information on all controls seems to have been affected more than the observations without information on all controls.

Using t-tests we found that the observation sample that provides information on all of the control variables have a mean GPA that is almost .5 grade points higher than the population mean. The left out observations had a lower mean GPA and also a smaller decrease in mean GPA after the grading scale change. The larger decrease in mean GPA in the regression that includes all control variables, regression five, could therefore be because of grade compression, which seems to affect programmes with higher grades more, as seen in the next section, 5.2.

5.2 Programmes with an above average GPA

As described in section 3.2, economic labour theory suggests that some students will be more motivated by a grading scale that provides more differentiation. When doing preliminary t-tests, a difference in the change of mean GPA before and after the policy change was found between programmes with high and low mean GPA. Therefore, to estimate whether there is any difference in how the new grading scale has affected high- and low achieving students, we ran separate regressions for programmes with higher and lower grades. The linear regression for examining higher achieving programmes, including controls, is

$$GPA_{it} | \text{mean GPA} > \text{average} \\ = \beta_0 + \beta_1 \text{after} + \beta_2 \text{educatedparents}_{it} + \beta_3 \text{women}_{it} + \beta_4 \text{immigrantbackground}_{it} + \varepsilon_{it}$$

The regression is similar to the one provided in section 5.1, with the difference that it only includes observations that have a mean GPA above the average. This means having a mean GPA above 15.6 grade points.

Table 5.2 Regression for programmes with higher than average GPA

Variables	<i>1. GPA</i>	<i>2. GPA</i>	<i>3. GPA</i>	<i>4. GPA</i>	<i>5. GPA</i>	<i>6. GPA</i>
<i>After</i>	-.4499***	-.4590***	-.4343***	-.4474***	-.4773 ***	-.5254***
	(.0334)	(.0338)	(.0321)	(.0452)	(.0451)	(.0447)
<i>Educated_ parents</i>		.0102***			.0071**	
		(.0024)			(.00363)	
<i>Women</i>			.0062***		.0065*	
			(.0023)		(.0033)	
<i>Immigrant_ background</i>				-.0181***	-.0159***	
				(.0035)	(.0035)	
Constant	15.8074***	15.1912***	14.3927** *	16.2441** *	15.3707** *	15.8938** *
	(.0055)	(.1516)	(.0917)	(.0688)	(.3568)	(.0087)
Observations	2,527	2,388	2,332	1,381	1,368	1368
R2	0.1072	0.1272	0.0844	0.2055	0.2132	0.1738
Standard error in parenthesis, clustered on schools. *** p<.01, ** p<.05, *p<0.1						

The regressions are presented in the same way as in table 5.1. The first column shows an overall regression without controls. The following three regressions are controlling for the share of students with highly educated parents, the share of women and the share of students with an immigrant background, respectively. Regression five shows the results when all of the controls were used. The last column shows the results of a regression that used the same sample as regression five, but without using any controls.

In all of the regressions, the policy seems to have had a negative impact on mean GPA. In regression five, the regression that includes all controls, the coefficient for the policy is about -.48 grade points, which means that the programmes with a higher than average GPA has had a decrease in their mean GPA by about 0.48 grade points. This can be compared to the overall regression in section 5.1, which had a slightly lower decrease of 0.47 grade points. This could be because of grade compression, since students who previously received a VG, but was near the lower boundary of VG, would now receive a D. When converting the grades into GPA using the scale mentioned

above, a VG would be worth 15 but a D would only be worth 12,5 and thus the mean GPA would have been lowered.

As in section 5.1, the sample that includes all the control variables has been affected more than when looking at the whole population, as can be seen in column six. As in section 5.1, this sample also has a higher mean GPA than the population. This also indicates that grade compression might have a bigger effect on programmes with higher GPAs.

5.3 Programmes with a below average GPA

To check the impact of the policy on programmes with a mean GPA below average, regressions that only included observations with a mean GPA below 15.6 grade points were run. The linear regression for this, including controls, is

$$GPA_{it} | \text{mean GPA} < \text{average} = \beta_0 + \beta_1 \text{after} + \beta_2 \text{educatedparents}_{it} + \beta_3 \text{women}_{it} + \beta_4 \text{immigrantbackground}_{it} + \varepsilon_{it}$$

The regression is similar to the one in section 5.1, the only difference being that it only includes observations with a mean GPA below 15.6 grade points.

Table 5.3 Regression for programmes with lower than average GPA

Variables	1. GPA	2. GPA	3. GPA	4. GPA	5. GPA	6. GPA
<i>After</i>	.0616	.0146	-.0958**	-.0764	-.1732***	-.2176***
	(.0378)	(.0396)	(.0402)	(.0586)	(.0620)	(.0620)
<i>Educated_</i> <i>parents</i>		.0085***			.0087	
		(.0027)			(.0056)	
<i>Women</i>			-.0004		.0047	
			(.0018)		(.0029)	
<i>Immigrant_</i> <i>background</i>				-.0194***	-.0148***	
				(.0037)	(.0043)	
Constant	13.0150**	12.750***	13.2676**	13.6796**	13.0671**	13.3416**
	*		*	*	*	*
	(.0096)	(.1141)	(.1029)	(.0968)	(.3825)	(.0218)

Observations	2,871	2,447	2,093	1,171	1,026	1026
R2	0.0021	0.0061	0.0060	0.0530	0.0665	0.0352
Standard error in parenthesis, clustered on schools. *** p<.01, ** p<.05, *p<0.1						

The table is presented in the same way as table 5.1 and 5.2. It starts with a regression without controls, then adds controls in the following columns. When looking at the results of regression one, which includes no controls, we get positive but insignificant coefficient for GPA after the implementation of the new grading scale. When controlling for the share of women we can see a decrease in GPA of .10 grade points that is significant at a 5% level, and when including all control variables we can see a decrease in GPA with .17 grade points that is significant at a 1% level. When only controlling for the share of students with highly educated parents or the share of students with an immigrant background we get insignificant results.

This indicates that both the programmes with higher than average GPA and programmes with lower than average GPA have a decreasing mean GPA after the implementation of the new grading scale. The policy seems to have had a negative effect on mean GPA, but the decreasing effect is lower for programmes with a lower than average mean GPA. An explanation for this could be the effect of decreasing grade compression.

Column six shows the results of a regression that used the same sample as the fifth column. As in sections 5.1-2, this sample has been affected more than when looking at the whole population, and as in sections 5.1-2, this sample has a higher mean GPA than the population. This also indicates that grade compression might affect programmes with higher grades.

5.4 Programmes with higher GPA

Literature suggest that high performing students will be more motivated by a stricter grading scale. To test whether the implementation of the new grading scale has had a different effect on higher achieving programmes, we ran regressions on the programmes that before the changed grading scale had a mean GPA above the 75th, 90th and 95th percentiles among all national programmes. These programmes have an average GPA above 15.6, 16.6 and 17.2 grade points respectively. Looking at the data we see that these programmes are often located in big city areas and are often natural science programmes. The schools are often private and the share of students with highly educated parents in the programmes is high.

Table 5.4 Regression for programmes with GPA above the 75th, 90th and 95th percentiles

	GPA above 75 %	GPA above 90 %	GPA above 95 %
Variables	<i>GPA</i>	<i>GPA</i>	<i>GPA</i>
<i>After</i>	-.4504*** (.0635)	-.4169*** (.1224)	-.0517 (.1024)
<i>Women</i>	.0041 (.0047)	.00631 (.0052)	.0027 (.0059)
<i>Immigrant_background</i>	.0040 (.0038)	.0101** (.0047)	.0010 (.0050)
<i>Educated_parents</i>	-.0096** (.0039)	-.0123* (.0067)	-.0241** (.0118)
Constant	16.2859*** (.5160)	16.4470*** (.5679)	17.8559*** (.6676)
Observations	721	310	145
R2	0.1912	0.1610	0.1008
Standard error in parenthesis, clustered on schools. *** p<.01, ** p<.05, *p<0.1			

The regressions show that the programmes with a GPA above the 75th percentile have decreased their average GPA with .45 grade points, and that programmes with a GPA above the 90th percentile have decreased their average GPA with .42 grade points. The programmes with a GPA above the 95th percentile have decreased their average GPA with .05 grade points, although this result is not significant, which might be because of the low number of observations. All regression in Table 5.4 includes all of the control variables.

Sections 5.2-3 showed that programmes with a higher than average mean GPA showed a bigger decrease in mean GPA after the grading scale change compared to programmes with a lower than average mean GPA. However, when looking at table 5.4 and the programmes with the highest grades the trend is the opposite – the higher grade the programme has, the lower is the decrease in mean GPA. This could indicate that there is a motivating effect that offsets the effect of decreasing grade compression when it comes to the programmes with the highest grades. Hence, it suggests that the implementation of the new grading scale could have had a motivational effect on higher

achieving students, since, if the mean GPA is above average, the decreasing effect of the new grading scale seems to be smaller the higher the grade is.

5.5 Control for time trend

As a way to control for a possible time trend and to see whether it is likely that the regression results are due to the grading scale change and not something that would have happened regardless of the policy change, we ran further regressions. The regressions are similar to the ones in sections 5.1-4, but use the mean GPA from year 2013 as the dependent variable, which means that we only use observations from before the new grading scale was implemented. Hence, in table 5.5 the variable “2013” represents the observations from year 2013, and the data includes observations from years 2010-2013.

Table 5.5 Control for time trend

	Overall	GPA above mean	GPA below mean	GPA above 75 %	GPA above 90 %	GPA above 95 %
Variables	<i>GPA</i>	<i>GPA</i>	<i>GPA</i>	<i>GPA</i>	<i>GPA</i>	<i>GPA</i>
<i>2013</i>	-.1268**	.1013	-.1493**	-.1721**	-.1442	-.1092
	(.0518)	(.0641)	(.0624)	(.0596)	(.0898)	(.0935)
<i>Educated_ parents</i>	.0109**	.0066	.0134*	.0089	.0107**	.0113**
	(.0043)	(.0049)	(.0080)	(.0056)	(.0050)	(.0053)
<i>Women</i>	(.0084)**	.0022	.0096**	.0048	.0134**	.0119*
	(.0033)	(.0043)	(.0037)	(.0046)	(.0058)	.0064
<i>Immigrant_ background</i>	-.0227***	-.0125	-.0129	-.0016	-.0016	-.0084*
	(.0052)	(.0054)*	(.0089)	(.0057)	(.0074)	(.0131)
Constant	14.2994***	15.5461***	12.5156***	15.7253***	15.7199***	16.2741***
	(.4046)	(.5394)	(.4795)	(.6510)	(.5771)	(.7951)
Observations	1768	1102	666	603	276	130
R2	0.0750	0.0372	0.0449	0.0361	0.0681	0.0910
Standard error in parenthesis, clustered on schools. *** p<.01, ** p<.05, *p<.1						

Looking at the results, a significant decrease in mean GPA can be found across all grade levels. However, the decrease is much smaller than the one found when using year 2014 as the dependent variable, which is the year that the policy was actually implemented. This could indicate that the estimates from the regressions in sections 5.1-4 are biased, and that there is a slight negative time trend from year to year. However, since the change in GPA is about four times as large when using 2014 as a as our dependent variable than when using 2013 as the dependent variable, one could still presume that the estimates are fairly accurate, although the magnitude might be misleading.

6. Discussion

6.1 Aggregated data

In this paper, data was gathered from the National Agency for Education's open database SiRiS. This database only shows data on programme level, not on individual student level, as explained above. Because of this, only the effect of the policy on programme level could be examined. Individual level data would have been preferable, since it then would have been possible to control directly for gender and socioeconomic background and the effect on the actual individual students could have been examined.

6.2 Missing data in the database SiRiS

In the database SiRiS, a figure that is based on fewer than ten students is shown as omitted. This is especially problematic for small programmes which will have a lot of missing data on controls. Due to this, the observation sample that provides information on all control variables will be biased. We can also see that the numbers of observations in the regressions are different when including controls variables. As seen in sections 5.1-3, the programmes that do not provide this information has a lower average mean GPA before the grading scale change and also a lower decrease in GPA after the changing of the grading scale. Access to all control data for all observations would likely have made our estimates more accurate.

6.3 Omitted variable bias

Due to limitations of the data set, only a few controls were used in the regressions and an omitted variable bias is probable. Additional controls such as class size, quality of teachers, GPAs from earlier education, location, and whether the school is private or public, among other things, are

probable to have an effect on mean GPA. However, using a fixed effects regression partly deals with this problem, since it only uses in-group variation. Assuming that, for example, a programme has approximately the same class size and the same teachers from year to year, the fact that one programme differs from another does not bias the estimates. However, since the class sizes, teachers' quality and other factors in the programmes might not actually be exactly the same from year to year, there might still be a bias.

6.4 The general school reform

Due to the reformation of the upper secondary school, the national programmes do not look exactly the same before and after the policy change, as mentioned above. Although we excluded the programmes that had changed the most, there might still be remaining problems. One example is the old Media Programme, which was removed in 2011. Which programmes would the students who formerly would have chosen the Media Programme choose now? Many of them might instead choose the Arts Programme, and looking at the mean GPA of the Media and the Arts Programme, we find that the mean GPA of the Media Programme is slightly lower. In 2011, when some potential Media Programme students now would choose the Arts Programme instead, that might have affected the mean GPA of the Arts Programme. These kinds of effects in the different programmes might also have biased the estimates in our regressions.

We also don't know if the general school reform and the changing of the formulation of the programmes have had other unobservable effects to the students' grades or the students' motivation. For example, if the formulation of the programmes and the course requirements look different after the school reform, this might affect the students' motivation and expectations on their education, which might have an impact on their results.

7. Conclusion

The aim of this thesis is to examine changes to the mean grade point average (GPA) after the grading scale change in the Swedish upper secondary school, and to examine whether there are any indications that the new grading scale has had an impact on students' motivation. The results can be of relevance to policy makers and the public since no prior quantitative studies on the effects of the new grading scale has been performed. Earlier studies on the new grading scale have examined students' and teachers' attitudes towards the new scale (National Agency for Education 2016).

The result show a decrease in overall programme mean GPA after the implementation of the new grading scale with .4688 grade points when controlling for the share of students with highly educated parents, the share of women and the share of students with immigrant background in every programme. As explained in section 5 and 6, this result is likely to be biased due to limitations of the dataset. However, it's still probable that the mean programme GPA has decreased, although more research would have to be done to be able to state an accurate magnitude. The result is in line with the theory on grade compression, which suggests that a broader grading scale will generate a lower GPA (Edgar 2014).

We also find indications that higher than average achieving programmes has had a smaller decrease in mean GPA the higher their grades are, which could indicate that the stricter grading scale and the additional grade levels has had a motivational effect on high achieving students. This would be in line with earlier research on the effects of the formation of the grading scale on students' motivation. Earlier studies suggest that the implementation of a stricter grading scale (Betts & Grogger 2003, Johnson & Beck 1988), as well as additional grade levels (Becker & Rosen 1992, Bressette 2002), may have a motivating impact for high ability students.

That students' GPA can accurately reflect the students' abilities is of high importance. Among other things it can be used as a signal of human capital for future employers, as explained in section 3.1. In order to help policy makers make informed and research based decision regarding grading policies, we hope that further research on the effects of the new grading scale will be performed. Further research with access to individual student observations would be interesting and could show more detailed effects. Some controls in our regressions were insignificant, which could indicate that they are not as relevant when using programme level data. Additional controls that would be interesting to include and are likely to have some effect are students' GPAs from earlier education, teacher quality and class size. Since theory also suggest that stricter grading could have a negative effect on the motivation for lower achieving student (Betts 1997), it is also important to closer examine the effects of the new grading scale for lower achieving students.

References

- Becker, W. E. and Rosen, S. (1992). The Learning Effect of Assessment and Evaluation in High School. *Economics of Education Review* 11(2): 107-118.
- Betts, J. and Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review* (22): 343-352.
- Betts, J. R. (1997). Do grading standards affect the incentive to learn? Discussion Paper 95-22. University of California, San Diego.
- Borjas, G. J. (2016). *Labor Economics*. 7th edition. New York: McGraw-Hill Education.
- Bressette, A. (2002). Arguments for Plus/Minus Grading: A Case Study. *Educational Research Quarterly* 25(3): 29-42.
- Children's Ombudsman [Barnombudsmannen] (2016). *En god utbildning*. Report by Children's Ombudsman.
- Ds 2008:13. *En ny betygsskala*. Stockholm: Utbildningsdepartementet.
- Edgar, L., et al. (2014). Student and Faculty Perceptions of Plus/Minus Grading and its Effect on Course Grade Point Averages. *College Student Journal* 48(1): 184-198.
- Elikai, F. and Schuhmann, P. W. (2010). An Examination of the Impact of Grading – Policies on Students' Achievement. *Issues in Accounting Education* 25(4): 677-693.
- Johnson, B. G. and Beck, H. P. (1988). Strict and Lenient Grading Scales: How Do They Affect the Performance of College Students with High and Low SAT Scores? *Teaching of Psychology* 15(3): 127-131.
- Miller, S. (1998). Shortcut: High School Grades as a Signal of Human Capital. *Educational Evaluation and Policy Analysis* 20(4): 299-311.
- Ministry of Education and Research (2004). *Könsskillnader i utbildningsresultatet – fakta, mönster och perspektiv*.
- National Agency for Education [Skolverket] (2016). *Utvärdering av den nya betygsskalan samt kunskapskravens utformning*.
- National Agency for Education [Skolverket], SiRis: Skolverkets Internetbaserade Resultat- och kvalitetsinformationssystem. The agency's online database on results and quality, <http://siris.skolverket.se> (accessed 22 April, 2016).
- Pattinson, E., et al. (2013). Is the Sky Falling? Grade Inflation and the Signalling Power of Grades. *Educational Researcher* 42(5): 259 - 265.
- Singleton, R. and Smith, E. (1978). Does grade inflation decrease the reliability of grades? *Journal of Educational Measurement* 15(1): 37-40.
- SOU 1992/93:250. *Ny läroplan och ett nytt betygssystem för gymnasieskolan, komvux, gymnasiesärskolan och särsvux*. Report by Läroplanskommittén.
- Spence, M. (1973). Job Market Signalling. *Quarterly Journal of Economics* 87(3): 355-374.