Stockholm School of Economics
Department of Economics
659 Degree project in Economics
Spring 2016

A for Effort: On Resolving Unfair Grading Practices by Reforming Swedish Upper Secondary School

David Berselius (23159) and Richard Sunnegårdh (22971)

Abstract: In 2011, upper secondary school in Sweden was reformed with the aim of reducing established differences in grading standards by imposing redesigned curricula and a new grading scale. This paper investigates the effect of this reform on grade inflation and evaluates its effectiveness in promoting fairness in grading between independent and public schools. We define grade inflation as the net difference between final course grades and national test scores. Using 2011–2014 panel data, we examine how the reform affects absolute levels of grade inflation and differences between school types. We perform the analyses both on subjects aggregately and separately, by regressing with school-specific fixed-effects. Our results indicate a significant increase in net grade inflation in total and in two out of three subjects. Furthermore, we find that the tendency of independent schools inflating grades more than public has decreased overall, but that differences have shifted differently across subjects. Definitive conclusions of the effectiveness of the reform are therefore difficult to make, but this paper shows that the it has had an effect on grade inflation and grading practices.

Keywords: grade inflation, grading standards, upper secondary school, criterion-referenced grading, reform
JEL Classification: I20, I24, I28

Supervisor: Federica Romei
Date submitted: May 15, 2016
Date examined: May 31, 2016
Discussants: Elin Boström, Elis Örjes
Examiner: Örjan Sjöberg

Acknowledgements

We would like to extend our sincerest gratitude to our supervisor Assistant professor Federica Romei for guidance and constructive feedback. We also direct many thanks to Luca Facchinello and Ylva Hedén Westerdahl whose advice has been crucial. Finally, we would like to thank our peers for their insightful critique and interest in our topic.

Contents

List of Figures

List of Tables

1 Introduction and purpose

For both society and the individual, education is fundamental for development and growth. Throughout history, grades have been used to assess educational achievement and knowledge, and to distinguish between students. When used properly, grades are a viable and efficient instrument, but if grading practices differ they may carry adverse effects. Thus, inflated grades have in recent years been a topic of discussion in Sweden and internationally, highlighting concerns about grades not reflecting *true* achievement, and how school types, i.e. public versus independent schools, grade differently. The increasing prevalence of these issues in Swedish upper secondary school was one main reason for imposing a reform in 2011, which brought about redesigned curricula and a new grading scale. In this paper, we use panel data on the relationship between national tests and final grades to evaluate the effect of this reform. How effectively can political interventions mitigate differences in grading standards?

Many studies have been conducted on the causes and consequences of grade inflation. Models from current research suggest that grade inflation undermines the signaling value of grades, removing one of their key uses (Yang and Yip 2003). Furthermore, lower grading standards harm both student achievement and quality of education (Bonnesrønning 2004; Figlio and Lucas 2004; Vlachos 2010). Upper secondary school has received relatively little attention in comparison to college-level, but there have been some influential papers in this field also. One major finding is that independent schools consistently practice more lenient grading standards (Wikström and Wikström 2005; Vlachos 2010; Nata *et al.* 2014), which is the focal point of this paper. Research also shows how school choice affects both grade inflation and access to higher education (Walsh 2010; Nata *et al.* 2014) and how informational asymmetries may further polarize society (Bunar 2010; Böhlmark *et al.* 2016).

It is of relevance to acknowledge the uniqueness of the Swedish school system, where grade-setting is entirely done locally and the central controls are limited. The grading is criterion-referenced and measures students on reaching certain criteria stated in national curricula in each course. More so than most countries, the Swedish university admissions rely heavily on the grade point average (GPA), amplifying the importance of grading practices. Independent schools were deregulated in 1992, introducing competition to the school system, and the number of these schools increased rapidly. Due to varying grading practices, and lagging student achievement a reform was passed in 2011, introducing a new grading scale, clearer curricula, and tighter central control. In 2016, an extensive report was released (Swedish National Agency for Educaton [SNAE] 2016b), surveying teachers and students on their views of the reform. They also performed statistical assessments of variations in grading practices across schools, finding an increasing spread overall. The report did however not discuss the effects on levels of grade inflation or the relative impact on independent and

public schools.

Considering previous research and in light of the reform, the purpose of this paper is to describe the reform's impact on absolute grade inflation and to evaluate its effectiveness in promoting fairness in grading practices between independent and public schools.

In order to do this examination we use 2011–2014 data from the SNAE (2016a) on the relationship between final course grades and national tests from each school unit[1] in every subject that has a national test. With the data, we define grade inflation as the net difference between the proportion of students receiving a higher course grade than on the national test and the proportion that receives a lower grade.

In this paper we analyze the impact of the reform in three steps. First, we compare absolute differences in grade inflation before and after the reform, in order to gain an understanding of its impact. Second, we compare the reform's impact on independent schools relative public, and can therefore isolate its effectiveness in alleviating differences in grading standards. Third, we perform the same analyses on the subjects English, Mathematics, and Swedish separately. Thus, we are also able to grasp the underlying drivers of the aggregate analyses.

Our results indicate that, after the reform, the grade inflation has increased aggregately, and that differences between independent and public schools have decreased. When studying the reform's impact on different subjects, we find rising levels of grade inflation in Mathematics and Swedish, and both increasing and decreasing differences between school types.

In conclusion, there is an increase in inflation, but that is not necessarily a result of lower standards. Grading is more uniform between independent and public schools overall, but as subjects shift in different directions, a final verdict is not easily made on the efficacy of the reform.

The paper is organized in the following manner: Section 2 defines our research question and states our hypotheses. In Section 3, we cover the background of the Swedish school system and previous research on grade inflation. Section 4 presents our data set and its biases, whereafter Section 5 describes the econometric methods used for the analyses and their application. Section 6 contains the empirical results, which we discuss in Section 7. Section 8 concludes the paper and identifies areas for future research and policy implications. Lastly, Section 9 summarizes the paper.

[1]SNAE statistics are presented per school unit—i.e. schools in most cases. Some large public schools are split into multiple school units. We use the terms *school* and *school unit* interchangeably.

2 Research question

We intend to contribute to the current state of knowledge and build onto previous research by answering the following research question:

Has the reform of 2011 impacted levels of grade inflation and resolved issues of differing grading standards between independent and public schools in Swedish upper secondary school?

As SNAE has not determined whether the reform has resulted in increased or decreased levels of grade inflation, we find it relevant to first understand its overall effects. Thereafter, we can investigate how independent schools, who have previously been shown to inflate grades more, have been affected differently than public. Therefore, we state the following hypotheses to answer our research question:

H_1: *The reform has had an effect on grade inflation*

H_2: *Independent school have been impacted differently from public schools*

Furthermore, we extend our scope of research to the effect of the reform on separate subjects. As curricula have not changed uniformly across subjects, we state two sub-hypotheses for each subject with a national tests; English, Mathematics, and Swedish:

$H_{subject,1}$: *The reform has had an effect on grade inflation in English/Mathematics/Swedish courses*

$H_{subject,2}$: *Independent school have been impacted differently from public schools in English/Mathematics/Swedish courses*

3 Background

In this section, we first go through how school choice in Swedish upper secondary school has developed. We then describe the criterion-based grading system and its characteristics. Next, we contextualize the reform and present its elements. Thereafter, we relate to previous research on grade inflation, its causes, and its consequences.

3.1 School choice in Sweden

Before 1992, the Swedish school system was a public monopoly with very few exceptions, with religious and boarding schools accounting for less than one percent of students. A legislation was passed that allowed for free school choice and for independent schools to be established. Control over the public schools shifted from federal to municipality level, and a system of one voucher per student came into effect as the means of funding. Initially, foundations and economic associations were establishing most of the independent schools, but over time private corporations have taken over their role. As of 2015, nearly half of Sweden's upper secondary schools are independently owned, with about 30% of the student population (SNAE 2014b, 2015, 2016a). Only three percent of students attend independent schools run by non-profit organizations (Timbro 2013). As there is a greater demand for students than there is supply, both public and independent schools are forced to compete for students in order to receive funding—before 1992 students were often assigned to the nearest school. This has created a market where students become customers and schools have to satisfy their needs. There is a wide variety of schools in both size and profile, where programs range from vocational training to university preparatory. Also, students apply to upper secondary school with their grades from compulsory school.

3.2 Criterion-based school system

In the beginning of the 1990s there was a large discussion taking place debating the grading system in the upper secondary school in Sweden. The final outcome was that in 1994, the Swedish school system was changed from a norm-based grading system along a normal (Gaussian) curve into a criterion-based assessment with distinctions between what constitutes a certain grade. One of the main advantages of this approach is that it provides a more individualized system where the teacher is be able to track the progress of each student more easily and increase the informative role of the grade (SOU 1992; Wikström 2005b).

The grades have three purposes: as a performance indicator and motivation on an individual level, as a general educational evaluation, and as a selection instrument for higher education (SNAE 2009). In a criterion-based system, the selection aspect is down-prioritized by the former two, as there are challenges to rank students with

a criterion-referenced instrument (Wikström 2005b). Theoretically, there is no limit to how many students that can have the same grade, thus limiting variability and the possibility of ranking. However, the system is argued to work as long as teachers are being fair and evaluate students equally across all schools, avoiding giving rise to systematically differing grading standards (Nata *et al.* 2014). These issues are amplified by the fact that Swedish university admissions rely solely on either the GPA or the SweSAT.[2] In other countries, universities oftentimes evaluate applicants on several different aspects combined. The implications of the Swedish system will be further developed in later sections.

Since 1994, the Swedish school system has become very unique, compared to the rest of the world, when decentralizing the responsibility of grading almost entirely to the individual schools—there are no systematic centralized controls. There are national tests (standardized tests) available in some of the main courses, but they are still being graded locally and only sporadically assessed centrally. Due to this lack of external control, Wikström and Wikström (2005) argue that there is room for grade inflation. SNAE pinpoints four different types of explanations for deviations between national test grade and course grade:

- The teachers interpret the goals and grading criteria differently
- The teachers give students grades that they do not deserve
- The course grade is evaluated on different grounds than just the centrally provided goals and criteria
- There are certain complementing tests/hand-ins made possible for students that fail the test, but can go on to complete the course with a passing grade (SNAE 2009)

Furthermore, grades have historically been a reliable indicator for future academic results—more so than the SweSAT—which makes it a useful selection instrument for higher education (Björklund *et al.* 2010).

3.3 School reform of 2011

Sweden, although among the best countries in the world from an educational point of view, experienced some concerns including that of lacking uniformity in grading standards, lower scores on international comparative studies, high drop-out rates, and relatively high youth unemployment rate in the beginning of the 21^{st} century. These issues brought about the school reform of 2011, for both upper secondary school and compulsory school, although only the former discussed is here (SNAE 2011).

[2] A standardized test battery that measures verbal and quantitative acumen, held biannually. Students may choose to apply to higher education with their scores.

The reform gave rise to several large changes, but we will focus on those that affected the issue of grade inflation. Two overarching goals were to make the school system more congruent, with fewer local courses and orientations, and provide clearer curricula and material for teachers and students. This was implemented along with a change from the four-step grading system (MVG, VG, G, IG) to a six-step system (A–F), which would provide more accurate grading with five instead of three passing grades for teachers to appoint students—before, the trenches were considered too wide. In the new system, the three grades A, C, and E, require students to reach all course criteria in order to obtain the final grade. At the same time, grades B and D do not have defined criteria, instead they are seen as intermediary grades for when students reach a majority of the criteria of the higher grade. Furthermore, from 2011 and onwards the national tests have been collected for central re-evaluation more frequently and systematically (SNAE 2011).

One year after the reform came into affect, a report was made covering teachers' sentiments on the consequences of the reform on both the students and themselves. The study shows that roughly three quarters of the teachers found that the new reform provided a fairer evaluation when grading and about 70% of them also believed that the new grading system provided more information about the students' academic level. They also find that having five different passing grades is useful for better precision. Even though they believe the new system to be more just, they do not find it easier to use; they understand the course material, criteria, and grade steps on their own, but find it difficult to integrate them in their entirety (SNAE 2014a).

In 2016, SNAE published an extensive report evaluating the implications of the new grading system and its criteria. Among other things, they study how the uniformity of grading standards has been affected. In line with the SNAE report from 2012, questionnaires find that teachers believe that they can grade more fairly with more grade steps. Relevant to our work, they also analyze the net differences between national test scores and course grades across the two systems—what we define as grade inflation. They study the variation of these differences between schools, which is an indicator of unfair grading. The variation has increased after the reform indicating that the differences in grading practices have increased overall, which is opposite to the intention of the reform (SNAE 2016b).

3.4 Grade inflation

Grade inflation has been a heavily discussed and researched topic in recent years. Many have expressed worries that too high grades result in inaccurate measures of academic performance. Most papers have focused on the phenomenon's presence in higher education, and economists have extensively mapped the characteristics of grade inflation in universities and colleges (Sabot *et al.* 1993; Wilson 1998; Yang and Yin 2003). In upper secondary school settings there have been fewer studies, but important research has been done by Walsh (2010), Nata *et al.* (2014), and Wikström

and Wikström (2005), and their findings have inspired much of our work.

3.4.1 Definition

The definition of grade inflation varies between studies. Yang and Yin (2003) distinguish between two common interpretations. The first, as some commentators argue, is rather a case of "grade compression," as when increasingly large proportions of classes receive the highest grades. This type of grade inflation results in a loss of information, where observers cannot differentiate exceptional students from average ones. The second analogizes to currency, in a sense that "bad grades drive out good grades"—a reference to Gresham's Law[3]—where higher grades will always be favored over lower ones. Therefore, if grade inflation were present in the school system as a whole, schools would punish its own students by not following suit. Both definitions capture a continuous increase in grades independent of student achievement. A third nuance of the concept focuses on the differential between individual students' grades and true ability in relation to other students. Here, development over time is less significant and schools' tendency to unfairly benefit its own students is the main concern. This is also the definition used in most literature at the upper secondary school level, and the one we will consider in this paper. It is worth noting that grade inflation in this sense is not undesirable *per se*, if practiced similarly between schools. Instead, it is systematic imbalances that may be harmful.

3.4.2 Access to higher education

One consequence of grade inflation that has been subject of research is that of access to higher education. In countries where grades are central in university acceptance, grading becomes an important factor in school choice since it may affect one's future academic studies. In a 2014 study, Nata *et al.* show that grade inflation among private and independent schools[4], which are generally populated by wealthier students, contribute to unequal access to higher education and impede social equity progress in Portugal. Though independent schools are free of charge in Sweden, socioeconomic factors do matter in school choice. Educated parents are often more aware of school quality and grading standards, and are more likely to influence schools in their favor. Similarly, students with high expectations from home will make a more educated choice of schooling. Depending on preferences, choice could therefore be based on a relatively better understanding of school quality or grading standards. Such examples of information asymmetries could in the long run result in further polarization in society (Bunar 2010; Söderström *et al.* 2010).

[3]When one type of money is overvalued and another type undervalued, only the overvalued money will be in circulation.

[4]In the case of Portugal, the distinction between private and independent schools is that the former take fees from their students. Neither are publicly owned.

3.4.3 Educational quality and signaling

Another concern about grade inflation is that of education quality and student achievement. From a knowledge perspective, research shows that strict grading standards generate higher achievement and, conversely, lenient grading harms performance (Ziomek and Svec 1997; Bonesrønning 2004; Figlio *et al.* 2004). In Sweden, this translates into a spiral of rising grades since 1994, even though the knowledge levels have been falling among students (SNAE 2009). Though dynamics of grading standards and achievement are difficult to assess by economic theory, it should be in the interest of policy-makers to improve student achievement, and educational systems should be designed to prevent inflation of grades. Historically, Sweden's criterion-based system has instead been proven lacking consistency of grading among schools and low uniformity in assessing students' understanding, which rather builds onto the inflation and ultimately undermines quality in education (Vlachos 2010).

A key concept in relation to our work and these concerns is that of signaling. In essence, signaling is to undertake certain activities mainly in order to affect others' interpretation of oneself. Education literature emphasizes two differing views on the role of education: capital formation and ability signaling. Traditionally, the main role of education is forming human capital, i.e. to learn things, which is also the view held by most teachers. On the other hand, students rather attend school to signal ability through grades. If grades truly measure underlying abilities, they provide a useful tool to employers and academic institutions to assess students, and for capable students to differentiate themselves (Spence 1973; Bedard 2001). When grades are systematically inflated and the cost of obtaining them decreases, they lose their signaling value as they contain less information (Waldow 2014). This effect has previously been modeled at the university level, and is framed as a free-riding problem. By inflating grades of mediocre students, schools allow them to "free-ride" on the reputation of high-performing students in order to secure jobs. Also, grade inflation is of "a contagious character: inflating schools mutually reinforce each other's practices" and, in the model, results in negative welfare-outcomes for students, firms, and universities (Yang and Yin 2003). In Swedish upper secondary school, it is reasonable to assume similar dynamics are present, especially because of the grades' absolute importance in university admissions.

3.4.4 Competition and grade inflation

Furthermore, competition as a driver of grade inflation has been one of the main branches of research in recent years. Pioneers on the subject are Wikström and Wikström (2005), who examine differences in grade inflation across Swedish municipalities with varying degrees of competition and presence of independent schools in 1997 by comparing student GPA and SweSAT scores. The study's main finding is the effect of a student attending an independently versus publicly run school, where the average

student receives significantly more inflated grades if the school is independent. The pattern has been consistent in more recent studies, however to a somewhat smaller extent. SNAE present the growing school market and differences in size as possible explanations (2015, 2014b, 2013, 2012). Though independent schools generally inflate grades more, competition on its own does not necessarily drive inflation in Sweden (Wikström and Wikström 2005; Vlachos 2010). Also in international settings, competition and marketization of education seems to be promoting inflation of grades. In Portugal, as mentioned above, fee-funded private schools are inflating grades (Nata *et al.* 2014). A study in the US shows that in school districts with high competition, schools respond by lowering grade cutoff points (Walsh 2010).

The introduction of competition in education is theoretically intended to give schools incentive to enhance quality and better match education to individual needs by directly connecting school funding with popularity. With enhanced school choice, the market of schools is obliged to adjust according to customer needs. Therefore, the informed customer, who is the parent or student, can strategically select schools to their own advantage both in terms of quality and leniency in grading (Walsh 2010). Welfare-maximizing schools should therefore adjust grading standards, if favored by customers, which would especially hold true for independently owned schools in a voucher setting, as revenues are directly associated with the number of students, i.e. popularity. When there is little central control over true learning outcomes, schools can even inflate at a low cost as there is small risk of "being detected cheating" (Wikström and Wikström 2005), and both schools and students are incentivized to favor lower grading standards.

Competition and its impact on grading does not necessarily work in the interests of students, especially those that are less informed. Asymmetric information between schools and students makes it difficult to determine quality in education. In contexts with similar asymmetries, such as health care, there are dire consequences for mistreatment. It is much more difficult to implement such controls in education, where quality is not easily defined. In Sweden, asymmetries are further emphasized as schools are rarely penalized for poor quality or differing grading standards, and educational goals are nuanced. By inflating grades, schools may create a competitive advantage rather than by increasing educational quality (Vlachos 2010).

4 Data

Every year, SNAE gathers data on all upper secondary schools in Sweden. Schools themselves declare a multitude of information on their students, grading and staff. Demographic data on students are also collected from Statistics Sweden, such as gender, parental education, and foreign background. Teacher education levels from the Swedish Higher Education Authority is also included. For our work, the data on national course test scores relative course grade is of highest interest, which has been collected since the fall semester of 2011. The data is presented as the percentage of test-taking students receiving lower, equal, or higher final course grade than received on the national test. National tests are taken for all Mathematics, Swedish, and English courses, and use the same scale as the final grade. SNAE publishes data sets on the course level in each school online on a yearly basis, and these will constitute the foundation of our analysis. In most other research, authors have used data from individuals. This results in some differences, but as our analysis targets differences in school behavior it should not impede our purposes. For the relevant years, 1,412 school units, out of which 483 are independent, are included. During the time period, some schools have been opened and some have been closed (SNAE 2016a).

There are some elements of the data from SNAE that cause biases. First, some schools do not report, or do so late, and are therefore not included in the data. Hence there may be a non-response selection bias, but its effects should be negligible. Second, SNAE does not report for tests taken by less than ten students at a school, and our analysis could therefore neglect the characteristics of very small schools or classes.

As a complementary data set, used for control, we collected yearly data on average admission grades from ninth grade for each school. SNAE could only provide admission records for 2013 and 2014 per program and school. For the remaining years, we contacted and gathered data from each of Sweden's 26 school districts. This data contains similar biases as the previously described data set. Data points are mainly missing due to some schools administering admission by themselves without reporting to the local school district. Descriptive statistics of our data is presented in Table 6 in Appendix A. Control variables are explained further in Section 5.4.

5 Method

In this section, we first motivate our overall method of analysis, how we construct the data, and define our dependent variable. Thereafter, we specify our first test and model how to test our first hypothesis, and then go through the same procedure for our second hypothesis and extended analysis. Last, we present how we apply the models on our data set.

To test our hypotheses, we regress grade inflation linearly with two different models. As there are many different variables affecting school behavior, it would be optimal to perform a controlled randomized test. In such a setting, schools or grade levels[5] would randomly be assigned to use the new grading system, while others would keep the old one as a control group. Hence, the test would provide an unbiased estimate of the system's impact on grade inflation in relation to true ability. In practice however, it would be very costly and unfair to evaluate students under different criteria just for an test. Instead, we choose to observe the gradual implementation of the reform as a natural test, in order to answer our research question.

The data used to examine the effects was collected from SNAE as well as from local admission units. The main data set consists of panel data over four years, from 2011 through 2014, resulting in 20,890 data points for grade inflation per year, school unit, grade level, and subject. We combined this data with other publicized SNAE data and our collected GPA, and the variables of interest were matched with the corresponding school unit and year. In cases where schools distinguish between courses taken in fall and spring semesters, data points are aggregated to the corresponding full school year. We also make the assumption that data points are randomly missing.

For our analysis, we define grade inflation as the net percentage of students receiving higher final course grades than in the national tests in each subject per grade level. Doing this, we subtract the number of students receiving lower grades from the students receiving higher, and divide by the total number of students taking that same course. To accurately analyze the true grade inflation, we would need to evaluate the true knowledge and ability of each student compared to his or her final grade in the respective course. In the best of worlds, this evaluation of true knowledge would have consisted of a nuanced assessment of the understanding of each subject. This assessment is difficult to make, and the national tests are the best proxy at hand. Thus, we will use the net percentage of deviations to assess how interpretation and application of criteria vary across schools.

5.1 Test 1: The reform's effect on grade inflation

Within our data, there are two separate groups: one under the new system and one under the old. This distinction is the basis of our analysis. In this first test, the

[5]To avoid confusion, we refer to what grade 1–3 a certain cohort is currently in as *grade level*.

purpose is to determine if the reform has had any significant impact on grade inflation on the population of schools as a whole.

As characteristics and students of individual schools are likely similar before and after the reform, we assume that the schools are homogenous over time. Adjusting for factors unique for the schools, we should then be able to estimate what effect the new grading system has had on grade inflation.

To test our first hypothesis, we perform a linear regression with grade inflation being the dependent variable. Our setup is multi-dimensional, and includes a school identification, school grade level, subject, and time. Using fixed-effects per school unit, we reduce the threat of unobserved heterogeneity from omitted variable bias. The threat of omitted variable bias is a common issue that most studies have to address, and it may pose great problems as to the reliability of the results if not accounted for. However, controlling with school fixed-effects should be even more robust than previous research where county or municipality fixed-effects has been used (Vlachos 2010). We propose the following model to explain differences in grade inflation:

$$Gradeinflation_{i,j,s,t} = \beta_{0,i,j,s,t} + \beta_1 Reform_{i,j,s,t} + \beta_2 Controls_{i,j,s,t} + \alpha_{i,j,s} + \epsilon_{i,j,s,t} \tag{1}$$

In the model, i is an index for school unit, j a school grade level identifier, ranging from 1 to 3, s identifies subject, and t is a time index. $\beta_{0,i,j,s}$ represents the fixed-effects of the school unit, grade level, and subject, and allows us to control for factors not included in our model. $\alpha_{i,j,s}$ is an unobserved time-invariant individual effect and $\epsilon_{i,j,s,t}$ is a time-variant error term. $Reform_{i,j,t}$ is a binary variable measuring what system the observation is under. For control variables, we include all variables in the Data section above, which we believe may affect grade inflation. The parameter of interest is β_1, which is the specific effect on the group affected by the new reform, and accounts for the average differences between those who were in the old system versus the new.

Since schools are likely to have similar characteristics over the years studied, serial correlation and heteroskedasticity must be accounted for as error terms could be correlated over time. Due to the relatively short observed time-period but large number of schools, we use both fixed-effects and clustered standard errors in our analysis in order to account for such biases (Wooldridge 2010).

5.2 Test 2: The reform's relative effect on independent schools

In the light of previous research, we intend to examine how independent schools, who historically have inflated grades more, have been affected by the reform. The test is constructed using a difference-in-differences method (DID). This method is commonly used to determine the effect of a treatment over time, where one group is impacted by a change, and the control group is not. The effect is captured by comparing

the average changes over time in the dependent variable between the groups. Our test does not include an untreated control group, since all students entering upper secondary school in 2011 are subjects to the treatment. Therefore, we have to assume that independent and public schools are affected differently by the treatment. This is reasonable to suspect, as a goal of the reform was to increase fairness in grading standards across all school types. Whether or not our first test points toward a positive or negative effect of the reform, our hypothesis rather reflects the differences between school types. In this aspect, a DID regression might help answering this part of our research question.

To test our hypothesis, we regress grade inflation linearly. See Test 1 for notations. We also assume that the parallel trends assumption holds. This means that, under the treatment, other factors affect the groups to the same extent, i.e. they are progressing in parallel over time. In our context, we would assume that the composition and surrounding factors of the schools remain stable as time progresses. This assumption is normally tested by observing trends before the treatment over longer periods. Since our data set contains limited such information, we support our assumption by referring to SNAE reports on previous years (2016b). This is a limitation of this part of the study and a rather broad assumption which may limit the degree to which we can draw inferences from these results—this will be further discussed in Section 7.2. Again, the panel is regressed with school unit fixed-effects and clustered standard errors. We use the following model to evaluate difference-in-differences:

$$\begin{aligned} Gradeinflation_{i,j,s,t} = \beta_{0,i,j,s,t} + \beta_1 Reform_{i,j,s,t} + \delta Reform_{i,j,s,t} \times Independent_i \\ + \beta_2 Controls_{i,j,s,t} + \alpha_{i,j,s} + \epsilon_{i,j,s,t} \end{aligned} \tag{2}$$

Again, $Reform_{i,j,s,t}$ is a binary variable for when the new system is in effect. To estimate the DID we also include another binary variable for whether a school is independent, $Independent_i$, which is interacted with the reform variable. The same control variables as in Test 1 are applied. Standard errors are clustered on the school level. The parameter relevant to test our hypothesis is δ, which measures the effect of the reform on independent schools relative public.

5.3 Extended analysis

To further develop our understanding of the effects of the reform, we extend our analysis to the specific subjects. As SNAE has established in previous research, grading practices vary greatly between the three subjects with national tests. Our two tests are designed to estimate the aggregate effects of the reform, but there could potentially be different dynamics involved for the different subjects. The new curricula have not necessarily changed uniformly across subjects. Therefore, we perform Test 1 and 2 on the subset for each subject respectively. This translates into performing

three additional regressions per hypothesis. We use the same model setups as before, see Equations 1 and 2, but where the subject indicator s is fixed per regression.

5.4 Application of the models

Our tests are intended to evaluate how our dependent variable—grade inflation defined as above—is affected by the reform, first on an absolute level and then with respect to school type. For the binary $Reform$ variable, we define a dummy variable that displays 1 for when the new system is in effect, and 0 for the old. Similarly, for the Test 2, we define another dummy for school type, the $Independent$ variable, as 1 for independent schools and 0 for public. The remainder of this section describes how we apply indexations and control variables to our regression analyses.

Grade level and year

As presented in Equations 1 and 2, we include indices for both grade level and year. These are treated as continuous in our regressions.

Grade level is important to identify, since previous research observes a clear escalation of grade inflation as grade level progresses (SNAE 2015, 2014b, 2013, 2012). Grade level also serves as a general separator between courses within subjects. In those instances where SNAE does not explicitly state grade level, an approximation is made.

As grade inflation previously has been found to increase over time, the year is also taken into account when performing the regressions.

Subject-specific factors

Due to the substantial discrepancies between Mathematics, English, and Swedish in grading standards, we also index for subject. Practically, we generate dummy variables $Mathematics$ and $Swedish$, so that observations display 1 for the subject in question and 0 for others. English is used as base value, since the grade inflation has historically been lowest in the subject's courses. For the extended analysis we examine the subjects individually.

School-specific factors

To control for factors specific to individual school units, we generate dummy variables for each one. The regression intercept is set to make the prediction calculated at means of independent variables equal to the mean value of the dependent variable. In total, we regress with respect to 1,293 dummies, which lay the foundation for our fixed-effects analysis of the panel. There is a multitude of school specific characteristics, difficult or impossible to fully account for, that could be associated with grade inflation. In Wikström and Wikström's (2005) work, they use data on individual

students, which allows for more detailed variables, such as parental income, home environment, and what program the student attends. As we assume student populations should remain relatively uniform, the fixed-effects would capture such time-invariant characteristics on the aggregate level that we study. Municipality specific factors, e.g. voucher amount and local policies, may therefore also indirectly be accounted for.

Control variables

When performing the regressions, we sequentially add control variables to the model. Some of these could be accounted for in the fixed-effects as they are practically time-invariant, but we choose to include them in the analysis to observe their average effects on grade inflation. Many of the control variables have been inspired by variables included in previous research by Wikström and Wikström (2005), Vlachos (2010), and SNAE (2015, 2014b, 2013, 2012), but we also account for some additional effects that we believe may impact the outcomes. In Table 1 we describe each control variable in detail and also indicate if they occur in previous research. Some of these are not exactly the same, but they can be seen as proxies for the same controls. Not controlling for these variables would be to neglect an extensive part of previous research.

Table 1: List of control variables

Variable	Description
9^{th} ***grade GPA***	The average GPA of students entering the school in a given year, ranging 0–320. As students entering 2013 and onwards use a new system, we normalize these scores to the previous years' scale. This should indicate the underlying proficiency of the average student. The measurement could be biased since grade inflation is prevalent also in compulsory school, but could provide explanatory value (Wikström and Wikström 2005; Vlachos 2010).
Female percentage	The percentage of female students attending the school. Values range 0–100. We include this variable since female students have historically been shown to receive more inflated grades (SNAE 2015, 2014b, 2013, 2012).
Independent percentage	The percentage of schools in the municipality that are independent in a specific year. Values range 0–100. This variable is intended to act as a proxy for competition from independent schools (Wikström and Wikström 2005; Vlachos 2010).
Entering students municipality	The number of students entering upper secondary school in the municipality for a specific year. This variable should help explain differences between differently sized municipalities (Wikström and Wikström 2005).
Schools per student	The number of schools in the municipality divided by the number of students in the municipality for a specific year. This variable should act as a proxy for competition overall (Vlachos 2010).
Students in school	The number of students in the observed school unit in a specific year. This should help explain differences between school sizes (SNAE 2015, 2014b, 2013, 2012).
Enter index	An indexed value of number of students entering upper secondary school in a municipality compared to the first observed year. The first year corresponds to the value 1. This variable should explain how student batch sizes locally affect grade inflation, disregarding municipality size. This has not been included in previous research, but may be important since batch sizes in the years studied are volatile.
Students per teacher	Measured on the individual school level on a yearly basis. The variable should serve as a proxy for school spending as well as teacher impact (Wikström and Wikström 2005).
Educated parents percentage	The percentage of students that have parents with higher education at the observed school unit. Values range 0–100. This should indicate the impact of socioeconomic background of the student population (Vlachos 2010).
Foreign background percentage	The percentage of a school's students that have foreign background. Values range 0–100. Also this should help explain the effect of background (Vlachos 2010).

6 Results

The following section contains the results of our tests and also illustrations of general patterns that the analyses reveal. We present the results from Test 1, Test 2, and the extended analysis in said order.

6.1 Test 1

To examine the effects of the reform on grade inflation, we perform a linear panel regression with clustered standard errors and fixed-effects on the school unit level. We add control variables sequentially and, with all controls in effect, 19,591 observations contain all data points. 1,293 school units have at least one data point, and these form our clusters from which standard deviations are estimated.

As can be seen in Table 2, the reform dummy variable displays a positive effect on grade inflation across schools with 99% significance. All else equal, the proportion of final grades being higher than on national tests increased by an average of nearly seven percentage points, depending on controls included. This supports our first hypothesis that the reform has indeed had an effect on grade inflation, and we can *reject* the null hypothesis that there was no change. This finding is interesting, and somewhat unanticipated, as the reform's stricter curricula, increased central control, and clearer grade cutoff points, would intuitively suggest smaller net deviations. This would infer that a higher percentage of students would receive a higher final course grade than national test grade compared to before. Alternatively, this means that fewer students receive lower grades—the net effect would be the same. In practice, the true effect is most likely a combination of both.

Table 2: Regression table: Test 1

	(1) Inflation	(2) Inflation	(3) Inflation	(4) Inflation	(5) Inflation	(6) Inflation	(7) Inflation	(8) Inflation	(9) Inflation	(10) Inflation
Year	0.00203	0.00223	0.00417**	0.00507***	0.00526***	0.00540***	0.00583***	0.00596***	0.00596***	0.00546***
	(1.14)	(1.23)	(2.19)	(2.63)	(2.68)	(2.72)	(2.89)	(2.94)	(2.94)	(2.65)
Grade level	0.0605***	0.0606***	0.0611***	0.0617***	0.0617***	0.0618***	0.0617***	0.0617***	0.0617***	0.0616***
	(23.54)	(23.37)	(23.45)	(23.53)	(23.56)	(23.56)	(23.52)	(23.30)	(23.32)	(23.32)
Mathematics	0.226***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***
	(44.33)	(43.93)	(43.80)	(43.80)	(43.80)	(43.79)	(43.77)	(43.69)	(43.70)	(43.70)
Swedish	0.143***	0.142***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***
	(34.20)	(33.60)	(33.55)	(33.54)	(33.57)	(33.57)	(33.54)	(33.32)	(33.31)	(33.31)
Reform	0.0650***	0.0639***	0.0656***	0.0675***	0.0677***	0.0679***	0.0680***	0.0683***	0.0684***	0.0683***
	(12.67)	(12.38)	(12.53)	(12.76)	(12.80)	(12.81)	(12.81)	(12.82)	(12.82)	(12.82)
9^{th} grade GPA	-0.000120	-0.000181	-0.000208	-0.000224	-0.000224	-0.000224	-0.000226	-0.000234	-0.000232	-0.000233
	(-0.81)	(-1.20)	(-1.38)	(-1.47)	(-1.48)	(-1.47)	(-1.48)	(-1.52)	(-1.50)	(-1.50)
Female percentage		0.000941***	0.000965***	0.000985***	0.000992***	0.00101***	0.00102***	0.00103***	0.00103***	0.00104***
		(2.85)	(2.95)	(3.02)	(3.05)	(3.09)	(3.12)	(3.16)	(3.18)	(3.20)
Independent percentage			0.000707***	0.000545**	0.000511**	0.000478**	0.000443*	0.000450*	0.000450*	0.000452*
			(3.39)	(2.51)	(2.29)	(2.10)	(1.94)	(1.94)	(1.94)	(1.95)
Entering students municipality				0.0000278***	0.0000275***	0.0000275***	0.0000242**	0.0000257***	0.0000257***	0.0000259***
				(2.97)	(2.93)	(2.93)	(2.52)	(2.65)	(2.66)	(2.67)
Schools per student					-0.826	-0.672	-0.431	-0.191	-0.184	-0.190
					(-0.53)	(-0.43)	(-0.28)	(-0.13)	(-0.12)	(-0.12)
Students in school						0.00000811	0.00000718	0.00000563	0.00000564	0.00000516
						(0.72)	(0.64)	(0.50)	(0.50)	(0.46)
Enter index							0.0300	0.0320	0.0317	0.0312
							(1.36)	(1.44)	(1.44)	(1.41)
Students per teacher								0.000995	0.00100	0.00101
								(1.21)	(1.22)	(1.23)
Educated parents percentage									-0.0000662	0.00000978
									(-0.14)	(0.02)
Foreign background percentage										0.000266
										(0.59)
_cons	-4.160	-4.595	-8.530**	-10.39***	-10.77***	-11.07***	-11.95***	-12.22***	-12.22***	-11.22***
	(-1.16)	(-1.26)	(-2.22)	(-2.68)	(-2.73)	(-2.77)	(-2.95)	(-2.99)	(-2.99)	(-2.71)
N	20307	19897	19792	19792	19792	19792	19778	19591	19591	19591
R^2	0.380	0.382	0.383	0.383	0.383	0.383	0.383	0.384	0.384	0.384
adj. R^2	0.338	0.339	0.339	0.340	0.340	0.340	0.339	0.340	0.340	0.340
F	424.8	358.3	314.5	279.7	252.9	229.8	210.2	191.8	179.0	167.2

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Since our measure of grade inflation is based on the percentage of higher course grades minus the percentage of lower course grades, we plot the distribution of the two before and after the reform respectively. This allows for a more granular interpretation of the shift toward higher grade inflation overall. Figure 1 indicates that there is no significant shift in pattern as a result of the reform, which is also corroborated by SNAE (2016b). However, the proportion of grades that deviate have undergone a pronounced increase in dispersion, both on the higher and lower end. In other words, the share of students that receive a different final grade than their national test suggests has grown. Before the reform, most schools only gave a minuscule proportion of students lower grades signified by the tall peak close to zero. After the reform, the left tail has become fatter which means it has become more common for schools to give lower grades than before. In comparison, schools tend to inflate a higher proportion of grades, and the reform has amplified that behavior. Hence, the reform has impacted grading in both directions, albeit the aggregate effect is skewed toward inflation rather than deflation, as the regression results indicate. This can also be seen in Figure 2, where the fraction of equal final grade and test grade shrinks from 76.5% to 66.8%.

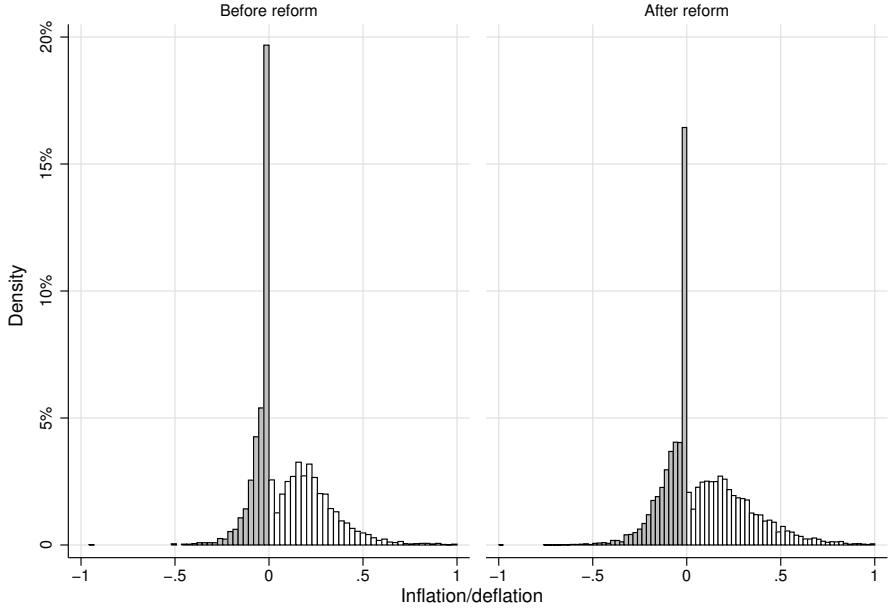

Figure 1: Distribution of proportions of students receiving higher (white bars) and lower (gray bars) final grades relative national test. One data point per school unit, subject, grade level, and year. The proportion of lower grades show a negative sign. Before and after the reform are presented separately. We observe similar patterns, but with increased dispersion after the reform. Nota bene, a high proportion of students receive the same final grade as on the national test, and that these are not included in the histograms.

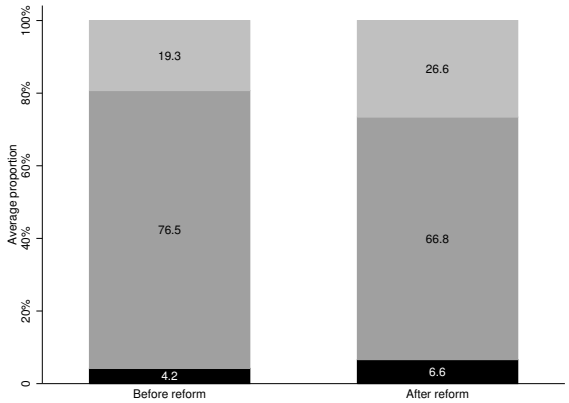

Figure 2: Average share of lower (black), equal (dark gray), and higher (light gray) final course grades relative national test over schools, before and after the reform. We observe increases in both higher and lower course grades, hence smaller proportions receiving equal. Averages are grade level-weighted.

In our regression results, the Grade level variable displays a highly significant positive coefficient of about six percentage points, meaning that grades are more inflated at higher grade levels. In Figure 3, we illustrate this trend by presenting the observations by grade level, where both mean increases and right tail becomes fatter for each grade level.

Also the subject has a major effect on grade inflation. In Swedish, and even more so in Mathematics, the net share of students receiving inflated grades is substantially higher than in English. The difference is, on average, 14 and 22 percentage points, respectively. In Figure 4, we plot distribution of grade inflation per subject and grade and find some distinct patterns. Firstly, there are very few national tests taken in English in the third grade. This is because the third course in English lacks a national test, and English does therefore not contain the same grade level effects as the other two subjects. That being said, grade inflation is also less common in English for the lower two grade levels. Secondly, lower final grades than on the national tests are very rare in Mathematics, and the right tail is much fatter than the other subjects.

Apart from Grade, Subject, and Reform, there are some other interesting results. Year seems to have a positive effect, though not statistically significant for all variations of the model. The percentage of female students is consistently related to higher inflation at a significant level, as well as percentage of independent schools in the municipality. Also the number of students entering upper secondary school in the municipality, controlling for differences in municipality sizes, increases grade inflation—larger municipalities tend to inflate grades more. The other controls do not show any significant results.

Lastly, we perform a control for whether or not level of inflation depends on the previous year's level. One could argue that schools that gave abnormally high

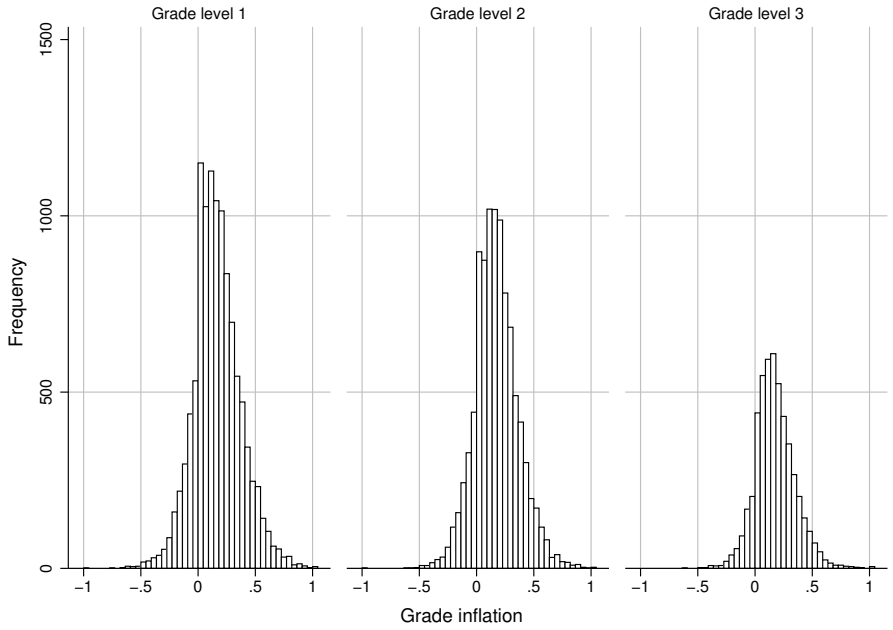

Figure 3: Grade 1–3 represent the distribution of net deviations in each grade level. One data point per school unit, subject, and year. Distributions are similar, but move toward more inflation as grade level progresses.

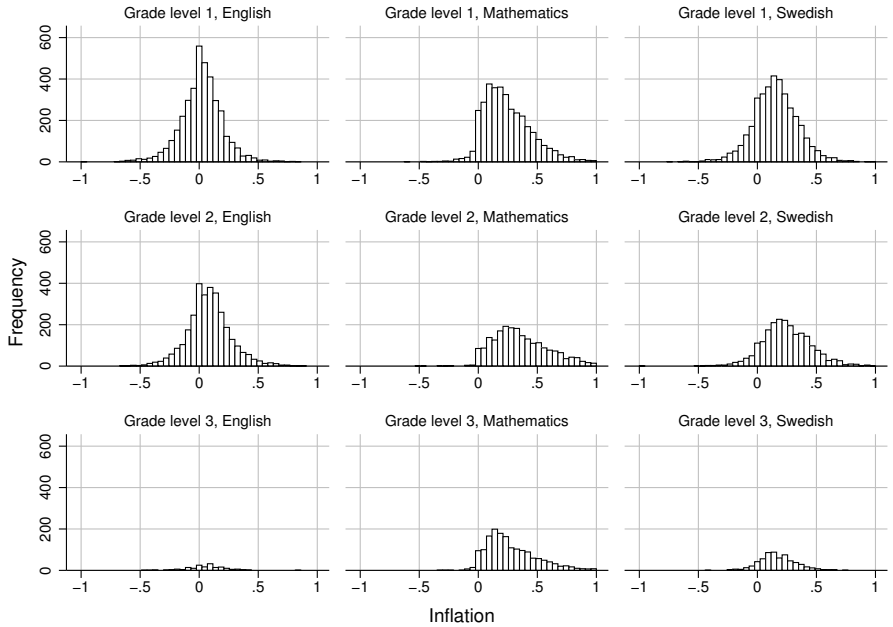

Figure 4: Distribution of grade inflation, per subject, grade level, school unit, and year. We find (1) few tests in English in higher grade levels, and (2) few negative net deviations in Mathematics

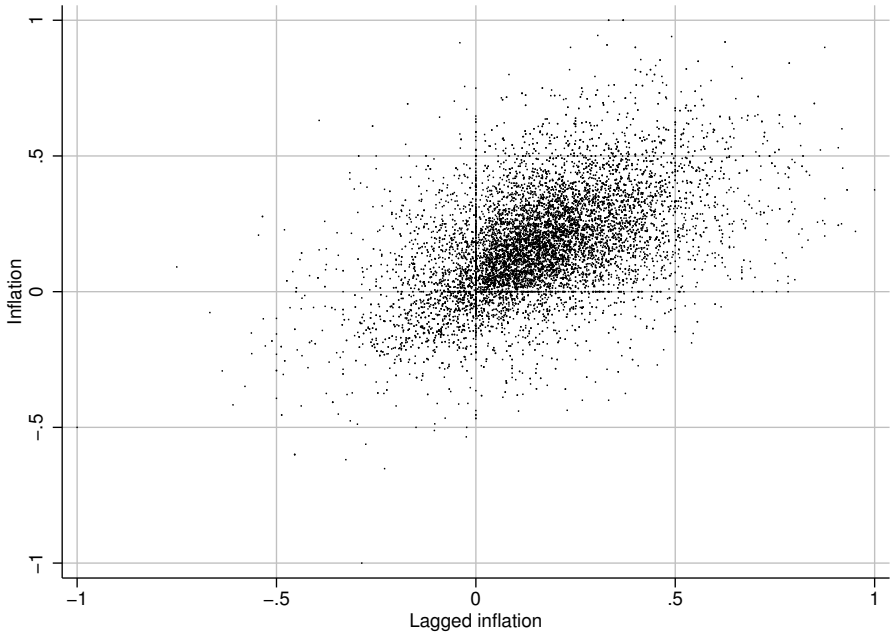

Figure 5: Relationship between grade inflation and its lagged value. Data points per subject, grade level, school unit, and year. Most observations in first quadrant, indicating consistent behavior over years.

proportions of its students inflated grades, would reverse its behavior the following year, and *vice versa*. As a visual representation, we plot inflation in year t-1 and inflation in year t. In Figure 5, we see that the lion's share of observations are in the first quadrant, suggesting that schools inflate quite consistently. Only the observations in the second and fourth quadrant represent schools that completely reversed their grading standards between years, and these are relatively few.

To test for potential multicollinearity, we construct a correlation matrix for our independent variables, see Figure 7 in Appendix B. When doing this robustness check, we find no evidence of any excessive levels of correlation. We find some interesting patterns, however, such as average GPA of student population correlates positively with both percentage of educated parents and female percentage.

6.2 Test 2

Before assessing the results of our second test, we investigate how, on an aggregate level, independent schools inflate grades in comparison to their public counterparts. In previous research on Swedish upper secondary school, independent schools have generally set more inflated grades (SNAE 2015, 2014b, 2013, 2012; Wikström and Wikström 2005; Vlachos 2010). We plot the distribution of higher, equal, and lower final grades of independent and public schools under each system in Figure 6. We find that independent schools indeed set more inflated grades, but that the discrepancy has

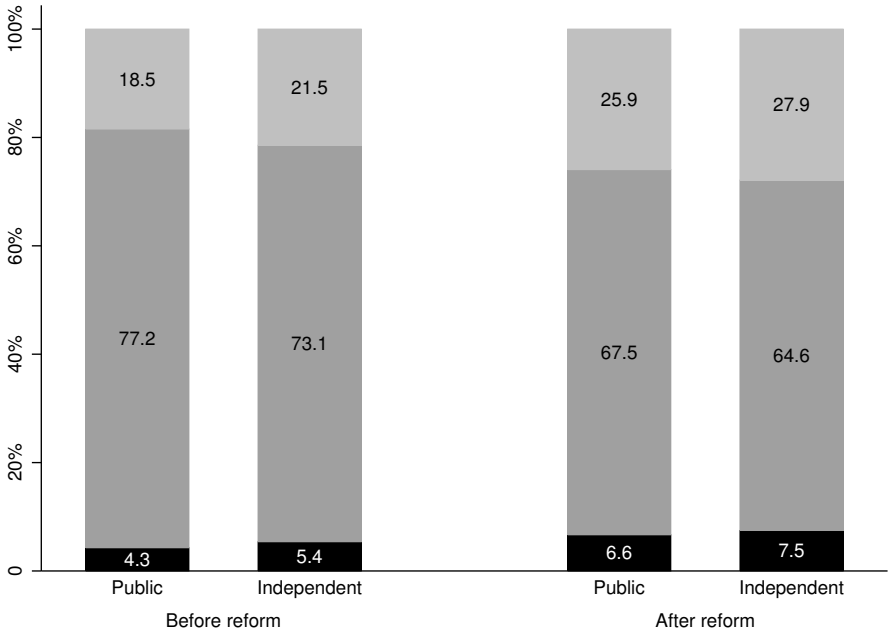

Figure 6: Average proportions of students receiving lower (black), equal (dark gray), and higher (light gray) final course grades relative national test in public and independent schools, before and after the reform. Averages are grade level-weighted. Inflation has increased for both, but to a smaller extent in independent schools.

decreased with the reform, also shown in Table 3. Note, however, that this analysis is based on averages and ignores effects of grade level, subject, individual schools, and control variables, and does not provide reliable answers to whether the reform has had an effect and, if so, to what extent. We therefore have to perform a more thorough analysis than using mere averages.

To test our second hypothesis, we perform a regression with clustered standard errors and fixed effects on the school unit level. We add control variables sequentially and, with all controls in effect, 19,591 observations contain all data points. 1,293 school units have at least one data point, and these form our clusters from which standard deviations are estimated.

Table 3: Difference-in-differences table

	Before reform	**After reform**	**Difference**
Independent	16.06%	20.43%	4.36%
Public	14.17%	19.28%	5.10%
Difference	1.89%	1.15%	**-0.74%**

Average inflation before and after the reform for independent and public schools. Grade inflation has increased with the reform for both school types, but to a smaller extent for independent schools.

Our results, that can be seen in Table 4, show statistically significant effects of our main point of interest; the interaction term of being an independent school after the reform. With 99% significance, we see that the difference in grade inflation between independent and public schools shrinks with the reform. This holds true for all variations of the our model and the decrease is, on average, between two and three percentage points. We can therefore *reject* our null hypothesis, that independent and public schools have not been impacted differently. Although there still is a difference in grading standards between independent and public schools, as shown in Figure 6, the reform has resulted in convergence when we perform the more thorough analysis. All parameters for control variables display the same sign as in Test 1. Multicollinearity between variables does not differ from the previous section, and the same analysis holds.

Table 4: Regression table: Test 2

	(1) Inflation	(2) Inflation	(3) Inflation	(4) Inflation	(5) Inflation	(6) Inflation	(7) Inflation	(8) Inflation	(9) Inflation	(10) Inflation
Year	0.00220	0.00240	0.00414**	0.00495**	0.00516***	0.00542***	0.00586***	0.00596***	0.00596***	0.00538***
	(1.24)	(1.33)	(2.17)	(2.57)	(2.63)	(2.73)	(2.91)	(2.94)	(2.94)	(2.61)
Gradelevel	0.0609***	0.0610***	0.0614***	0.0619***	0.0619***	0.0620***	0.0620***	0.0619***	0.0619***	0.0619***
	(23.60)	(23.42)	(23.49)	(23.55)	(23.58)	(23.58)	(23.55)	(23.31)	(23.34)	(23.33)
Mathematics	0.226***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***	0.227***
	(44.32)	(43.92)	(43.79)	(43.80)	(43.80)	(43.78)	(43.76)	(43.68)	(43.69)	(43.69)
Swedish	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***	0.143***
	(34.23)	(33.63)	(33.57)	(33.56)	(33.60)	(33.61)	(33.57)	(33.35)	(33.35)	(33.35)
Reform	0.0767***	0.0753***	0.0758***	0.0764***	0.0767***	0.0779***	0.0781***	0.0776***	0.0777***	0.0778***
	(13.23)	(12.91)	(12.86)	(12.93)	(13.04)	(13.08)	(13.08)	(12.95)	(12.94)	(12.96)
Independent*Reform	-0.0291***	-0.0279***	-0.0249***	-0.0224***	-0.0226***	-0.0247***	-0.0248***	-0.0231***	-0.0231***	-0.0234***
	(-3.54)	(-3.40)	(-2.99)	(-2.68)	(-2.72)	(-2.91)	(-2.92)	(-2.71)	(-2.72)	(-2.75)
9^{th} grade GPA	-0.000121	-0.000183	-0.000207	-0.000221	-0.000222	-0.000220	-0.000222	-0.000231	-0.000228	-0.000229
	(-0.82)	(-1.21)	(-1.37)	(-1.45)	(-1.46)	(-1.45)	(-1.46)	(-1.50)	(-1.47)	(-1.47)
Female percentage		0.000946***	0.000967***	0.000985***	0.000992***	0.00102***	0.00103***	0.00105***	0.00105***	0.00106***
		(2.85)	(2.95)	(3.00)	(3.04)	(3.12)	(3.15)	(3.20)	(3.22)	(3.25)
Independent percentage			0.000644***	0.000504**	0.000465**	0.000403*	0.000367	0.000376	0.000376	0.000377
			(3.06)	(2.32)	(2.07)	(1.76)	(1.59)	(1.61)	(1.61)	(1.62)
Entering students municipality				0.0000250***	0.0000247***	0.0000245***	0.0000211**	0.0000227**	0.0000227**	0.0000229**
				(2.67)	(2.62)	(2.59)	(2.18)	(2.34)	(2.34)	(2.35)
Schools per student					-0.948	-0.684	-0.438	-0.195	-0.185	-0.192
					(-0.61)	(-0.44)	(-0.29)	(-0.13)	(-0.12)	(-0.13)
Students in school						0.0000144	0.0000135	0.0000116	0.0000116	0.0000111
						(1.25)	(1.17)	(1.00)	(1.00)	(0.96)
Enter index							0.0306	0.0327	0.0324	0.0317
							(1.39)	(1.48)	(1.47)	(1.44)
Students per teacher								0.000889	0.000902	0.000903
								(1.09)	(1.10)	(1.10)
Educated parents percentage									-0.0000974	-0.0000102
									(-0.20)	(-0.02)
Foreign background percentage										0.000307
										(0.68)
_cons	-4.505	-4.933	-8.461**	-10.14***	-10.57***	-11.09***	-11.99***	-12.21***	-12.21***	-11.06***
	(-1.26)	(-1.36)	(-2.21)	(-2.61)	(-2.68)	(-2.78)	(-2.96)	(-3.00)	(-2.99)	(-2.67)
N	20307	19897	19792	19792	19792	19792	19778	19591	19591	19591
R^2	0.381	0.383	0.383	0.383	0.383	0.383	0.383	0.384	0.384	0.384
adj. R^2	0.338	0.339	0.340	0.340	0.340	0.340	0.340	0.340	0.340	0.340
F	363.7	313.4	279.6	251.7	230.0	210.7	194.0	178.1	167.1	156.8

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.3 Extended analysis

Our initial two tests reveal significant differences in grade inflation between subjects. In light of this finding, which is also supported by previous research (SNAE 2015, 2014b, 2013, 2012), we plot the average proportion of students receiving higher, equal, and lower final grades than on the national tests in each subject, before and after the reform. Figures 7–9 in Appendix C show that the reform has affected grade inflation in Mathematics and Swedish courses to a higher extent than in English. The averages also suggest that, while independent schools on an aggregate level inflate more, inflation in Swedish courses is actually higher in public schools than in independent—this difference appears to have increased. For both patterns, however, more thorough analysis is required.

Firstly, we regress grade inflation per subject with clustered standard errors and school unit fixed effects (as in Test 1). We obtain 6,593, 6,917, and 6,081 observations in English, Mathematics, and Swedish respectively. The results are presented in the first three columns of Table 5. The parameter of the $Reform$ variable displays highly significant positive coefficients for Mathematics (10 percentage points) and Swedish (12 percentage points), and we *reject* our null hypothesis for these two subjects. In English, the coefficient is also positive, but is *not* significantly different from zero. This is in line with the patterns observed in Figures 7–9.

Interestingly, schools with higher percentages of female students inflate more in English and Mathematics, but not at all in Swedish courses. Higher 9^{th} grade GPA seems to predict significantly less inflation in Mathematics, but more inflation in Swedish. Large municipality sizes and high proportions of independent schools also drive grade inflation in Mathematics. Increasing student batch sizes, quite surprisingly, correlate positively with grade inflation for English courses—increasing within-municipality student population sizes seem to drive inflation.

Secondly, using the same methodology as in Test 2, we perform a regression that includes the interaction between $Independent$ and $Reform$ for each subject. The results are presented in the last three columns of Table 5. For all subjects, though *not* statistically significant for English, the interaction terms $Independent{*}Reform$ show that the reform has had a negative effect (2.8 percentage points for Mathematics and 3.6 percentage points for English) on grade inflation in independent relative public schools. We therefore *reject* our null hypothesis for Mathematics and Swedish. In Mathematics, this points toward more coherent grading standards after the reform, but only at a ten percent significance level. In Swedish, where public schools actually tend to inflate slightly more, the reform has had the opposite effect, as the difference between independent and public schools disperses significantly.

Table 5: Regression table: Extended analysis

	(English) Inflation	(Mathematics) Inflation	(Swedish) Inflation	(English) Inflation	(Mathematics) Inflation	(Swedish) Inflation
Year	-0.00356	0.00136	0.0150***	-0.00359	0.00128	0.0145***
	(-1.12)	(0.36)	(4.05)	(-1.13)	(0.34)	(3.92)
Grade level	0.0617***	0.0637***	0.0964***	0.0620***	0.0637***	0.0979***
	(14.38)	(16.67)	(15.71)	(14.44)	(16.67)	(15.87)
Reform	0.00921	0.0996***	0.119***	0.0189**	0.111***	0.134***
	(1.10)	(11.07)	(10.45)	(1.97)	(10.51)	(11.01)
Independent*Reform				-0.0224	-0.0278*	-0.0362***
				(-1.63)	(-1.75)	(-2.66)
9^{th} grade GPA	0.0000416	-0.00102***	0.000542*	0.0000447	-0.00101***	0.000557*
	(0.18)	(-3.65)	(1.87)	(0.19)	(-3.62)	(1.92)
Female percentage	0.00168***	0.00122**	0.0000998	0.00170***	0.00124**	0.000126
	(3.33)	(2.22)	(0.18)	(3.35)	(2.24)	(0.22)
Independent percentage	0.000444	0.00119***	-0.000332	0.000380	0.00109**	-0.000459
	(1.25)	(2.74)	(-0.82)	(1.07)	(2.48)	(-1.12)
Entering students municipality	0.0000240	0.0000461***	0.00000890	0.0000214	0.0000421**	0.00000498
	(1.45)	(2.78)	(0.55)	(1.28)	(2.55)	(0.30)
Schools per student	1.784	-1.246	-1.077	1.766	-1.227	-1.100
	(0.63)	(-0.42)	(-0.24)	(0.63)	(-0.41)	(-0.24)
Students in school	0.0000100	-0.00000812	0.00000779	0.0000155	0.000000138	0.0000159
	(0.74)	(-0.38)	(0.38)	(1.11)	(0.01)	(0.76)
Enter index	0.0732**	0.0192	0.0125	0.0741**	0.0207	0.0119
	(2.23)	(0.46)	(0.27)	(2.26)	(0.50)	(0.26)
Students per teacher	0.00134	0.00141	-0.00111	0.00126	0.00126	-0.00125
	(1.15)	(0.92)	(-0.75)	(1.07)	(0.82)	(-0.84)
Educated parents percentage	0.000222	-0.00114	0.000413	0.000209	-0.00115	0.000362
	(0.28)	(-1.16)	(0.49)	(0.26)	(-1.18)	(0.43)
Foreign background percentage	-0.000255	-0.000113	0.000713	-0.000229	-0.0000600	0.000788
	(-0.37)	(-0.14)	(0.96)	(-0.33)	(-0.07)	(1.06)
_cons	6.838	-2.619	-30.30***	6.905	-2.442	-29.48***
	(1.07)	(-0.35)	(-4.08)	(1.08)	(-0.32)	(-3.95)
N	6593	6917	6081	6593	6917	6081
R^2	0.496	0.456	0.425	0.497	0.457	0.426
adj. R^2	0.376	0.336	0.273	0.376	0.337	0.274
F	22.92	24.66	32.38	21.54	23.17	30.54

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7 Discussion

In the following section, we first present our findings and discuss their explanations and implications. We then discuss the limitations of our research, and finally its external validity.

7.1 Findings

The reform's effect on grade inflation

By analyzing the implementation of the Swedish upper secondary school reform of 2011, we find evidence in line with our first hypothesis by concluding a *significant increase* in grade inflation as a result of the reform. Our conclusion holds on an aggregate level, and for courses within the subjects Mathematics and Swedish. We are *unable* to detect any significant differences for courses in English. The general patterns in grade inflation between subjects are similar to what has been published by SNAE (2012, 2013, 2014b, 2015). While SNAE shows that the variation in net deviations has increased (2016b), we add onto the knowledge by proving that also absolute levels have risen.

The reform's effect on the distribution of inflated and deflated grades can be attributed to several possible factors. Firstly, the reform entails two additional passing grade levels. The increase in possible outcomes would logically result in more deviations from the national tests, which is also confirmed by SNAE (2012, 2013, 2014b, 2015). Practically, the cutoff points become more concentrated, lowering the threshold for setting a different final grade. This effect would explain the patterns in Figures 1–2, 7–9, where the proportion of students receiving the same grade has decreased. The new grading scale could also reduce the "ceiling effect" that previous studies have found. As fewer students receive A on the national tests than received MVG before the reform, there would be more room for inflation. The percentage of national tests given the highest grade has dropped from about 5–10% to 1% after the reform (SNAE 2012, 2013). Secondly, as mentioned in the background section, teachers find the new grading system more difficult to implement. Said problem could be part of a transition phase, as teachers have to adapt and accustom themselves to the new system, but could also be a sign of inherent ambiguity in the new curricula (SNAE 2015). Thirdly, the new system could allow for a more complete evaluation of students' performance over the whole course. The relative contribution of these factors is difficult to determine, nevertheless the analysis remains that there has been a significant increase in grade inflation on average with the reform.

Our extended analysis suggests that the increase in grade inflation with the reform is concentrated within Mathematics and Swedish courses, while we cannot find a significant increase for English. One *could* infer that Mathematics and Swedish have been affected more by the factors mentioned above, but our analysis does not

prove any causality. The difference could just as well be a consequence of the varying construction of the subjects' curricula. As for the absolute differences in grade inflation levels between subjects, some explanatory value comes from the large portion of students receiving a failing grade—F or IG—on the national test in Mathematics. As there is no lower grade, only equal and higher final grades are possible. At the same time teachers are often more lenient in order to pass the student in these situtations, resulting in more inflation. It is not uncommon for 20% of all students taking courses in Mathematics to receive a failing grade on the national test, in comparison to Swedish and English where that figure usually is below 5% (2012, 2013, 2014b, 2015). This is supported by average admission GPA from 9^{th} grade correlating negatively with inflation in Mathematics only, suggesting that schools with lower admission scores inflate more.

With regards to grade inflation increasing with grade level in total and subject-wise, we find several possible explanations corroborated by SNAE (2012, 2013, 2014b, 2015). Students taking courses with national tests late in upper secondary school generally already have a relation with their teacher, which could incorporate earlier merits in grade-setting. Teachers may also take into account that these students tend to be studious and attending university-preparatory programs—the stakes are higher as each course grade may directly affect students' future academic careers. Conversely, late courses are often considered more difficult, and teachers may pass poor students even though they fail the national test. Also, a greater number of students take courses in the first year than the second and third, so teachers have more experience in grading these courses. Together, along with other possible factors, these contribute to the grade level effect that is apparent in our results.

We also observe some additional findings that support previous research. Firstly, the year has a statistically significant positive effect on grade inflation, just as Wikström (2005b) finds, suggesting a gradual progression over time. However, when observing the individual subjects, only Swedish courses display significance. Secondly, in accordance with SNAE (2012, 2013, 2014b, 2015), and Wikström and Wikström (2005), high percentages of female students predict higher grade inflation. For this trend, an explanation still remains to be found. Thirdly, competition, expressed through the percentage of independent schools, seems to drive grade inflation, which to some extent contradicts Vlachos (2010) and Wikström and Wikström (2005), who find no evidence for competition having an effect. Lastly, on an aggregate level, as well as for Mathematics separately, inflation increases with municipality size. This finding is in line with Wikström (2005a).

The reform's relative effect on independent schools

In our second test, we test our second hypothesis if independent and public schools have been affected differently by the reform. We first confirm SNAE (2012, 2013, 2014b, 2015), Vlachos (2010), Nata *et al.* (2014), and Wikström and Wikström's

(2005) findings that independent schools inflate grades more in total. This pattern is consistent also in Mathematics and English, but in Swedish courses, we find that public schools actually inflate somewhat more. Our analysis does not provide insight as to what might be the underlying reasons for this.

We find with *statistical significance* that the reform has reduced the gap in grade inflation between independent and public schools on an aggregate level. This finding alone would suggest that grading standards have become more homogeneous between public and independent schools with the reform. However, when analyzing the subjects separately, we see that independent and public schools have been impacted differently. In Mathematics, where independent schools generally inflate more, we find with 90% significance that the difference has shrunk. On the contrary, public schools inflated grades in Swedish courses more prior to the reform, and this discrepancy has grown larger with high statistical significance. We are *unable* to show any significant change in English courses, even though the coefficient displays a negative sign. Within subjects, grading standards have therefore not necessarily become more homogeneous overall after the reform. Nevertheless, we can conclude that independent schools now grade inflate relatively less. Whether we should attribute this shift to increased central control, the clearer curricula, or some other implication of the reform, is difficult to determine—yet it exhibits that the reform has had an effect. The degree to which we can state this with certainty is held back by the assumption of parallel trends, which is discussed below.

Putting these findings in an economic context, it appears as the new curricula and increased central control have restrained the incentives to unfairly inflate grades. We contribute this to two mechanisms. First, increased monitoring cause schools to be more cautious in altering grading standards. The mere awareness of the risk of being "detected cheating" reduces the incentive to do so. Second, schools that still do practice unfair grading standards are more likely to be penalized and forced to adjust accordingly. The extent to which schools can conceal lower educational quality by inflating grades is also likely to have been reduced thanks to tighter control. However, the mixed results when analyzing the subjects make inference problematic. Considering the issues regarding access to higher education and informational asymmetries between students, the difference between choosing an independent and a public school has become less emphasized in terms of grading leniency. Yet there are substantial differences between individual schools that may create unfair advantages for better informed students.

7.2 Limitations

When considering the broader applicability and reliability of our results, there are some limitations that should be considered.

Firstly, one of our largest limitations is assuming parallel trends in Test 2, which undermines some of the potential to draw conclusions about it. What we assume is

that if there had been no reform, the difference of independent and public schools' grade inflation had remained the same. Such an assumption is often robustness checked over longer periods in time to make sure the parallel trends hold true. Unfortunately, our data set does not allow us to perform this check and we have had to refer back to previous reports (SNAE 2016b). This undoubtedly puts a restrain on the DiD results, and this fact should be kept in mind when performing the analysis.

Secondly, teachers correct and grade national tests locally and also set course grades. Just as final grades could be set unfairly, national tests grades could be assessed using different grading standards. Subsequently, there may be more grade inflation, by our definition, than data would indicate. This concern is verified by reports done by the Swedish Schools Inspectorate (2013). One could argue that standardized tests with centralized grading would help reduce such outcomes, but studies have shown how results are manipulated by schools and teachers nevertheless (Jacob and Levitt 2003). In other courses than Mathematics, Swedish, and English, the lack of standardized testing makes it difficult to capture potentially inflated grades. Average course grades in courses without national tests have, in fact, experienced a faster growth in recent years (Vlachos 2010). Even without this information, we assume that the available data is indicative of relative grade inflation among schools. It should also be noted that national tests do not necessarily take into account all grading criteria that the final grades do—final grades are based on a more cumulative assessment of the intended learning outcomes than just the test score. Though our measure of grade inflation is not optimal, it should be able to reveal systematic differences in grading standards. Vlachos's use of the same measure supports its validity.

Thirdly, our panel only stretches over four years of data, which gives rise to some limitations. The data set is weighted toward observations after the reform, due to SNAE not collecting national test scores consistently before 2011. Data over longer periods of time would have improved the robustness of our results and ability to observe trends over time. However, thanks to our large number of data points every year, a majority of our results are significant with p-values below 0.01 and point in the same directions.

Fourthly, we do not distinguish between non-profit and profit-driven independent schools. The case may be that non-profit schools tend to inflate differently, but since the total number of students attending non-profit independent schools only amount to three percent of the population the analysis should not be affected substantially (Timbro 2013). Also, there is a small number of public schools that operate as if independent. SNAE does not distinguish between these in the data set, and potential differences are not accounted for in our analysis.

Lastly, our analysis only controls *if* schools inflate, not *how* they do so. The data set contains information on what proportion of students receive deviating final grades, without specifying between which grade steps deviations occur. Such information

could reveal additional dynamics of grade inflation, but should not change our results. Instead, it could help explain why certain patterns appear.

7.3 External validity

Before discussing in what contexts our findings may be valuable, we must first understand the fundamental differences between Swedish upper secondary school and that of other countries. In the following paragraphs we put the Swedish system into an international context with regard to the role of grades, curricula design, external control, and school choice.

Firstly, the stakes on grades from upper secondary school are higher in Sweden than in many other countries. In the United States and United Kingdom, for instance, external evaluations are the main decision criteria for admission to higher education. Due to the SweSAT, however, Sweden is more comparable to such countries than other Scandinavian countries, where no such alternative route to higher education exists (SNAE 2016b).

Secondly, Swedish curricula are rather unique internationally. In recent decades, most countries outside of Germany, Iceland, Denmark, and Norway, where more traditional goal-oriented curricula are in use, have implemented "standards-based" curricula. These do not only specify what abilities the student should show, but also the quality of those abilities. In Norway, Iceland, and Denmark, they describe in general terms what each grade step constitutes, while German curricula connect steps tightly to certain achievements. In contrast, Swedish curricula *abstractly* describe the quality required for the different grade steps. This could make it difficult to relate grading practices in Sweden to that of other countries (SNAE 2016b).

Thirdly, the role of external control varies greatly between countries. In the United States, the United Kingdom, Canada, China, and South Korea, for instance, national standardized testing are used as a primary tool for assessment of schools and students. Denmark, Finland, Norway, Iceland, and Germany have a tradition of *censors*, whose task is to monitor teachers. These may overrule local grading. Similar practices are used in Australia, where highly regarded teachers *moderate* their peers. Again, Sweden is quite unique, having national tests but relying on local teachers for final judgement. Therefore, the issue of equal grading standards is amplified in Sweden, and becomes a more important topic of discussion than elsewhere (SNAE 2016b).

Fourthly, not all countries have the same extent of school choice as Sweden. In many countries, independent schools take fees, which makes socioeconomic factors more prevalent than in Sweden. The voucher system in Sweden does, however, have some comparable counterpart internationally. Bettinger (2011) argues that the Netherlands has a similar system as Sweden, where public and independent schools compete for students. India, Chile, Belize, Japan, Canada, Colombia, and Poland, among others, also have voucher programs in use, though not directly translatable to the Swedish system.

Our work shows that criterion-based grading systems may give rise to differences in grading standards, but that school behavior can be changed through political intervention. By introducing a clearer curricula and stricter control, this has been possible in Sweden, and could potentially be applied elsewhere. However, our results also show difficulties in calibrating a new system, and that many factors should be considered. Another cost of reconstructing curricula is transition periods, where teachers and students have to adjust to new circumstances (SNAE 2016b). In terms of external validity, we use Sweden as an example in how to approach unfairness in grading standards in the context of criterion-based grading systems. Due to the uniqueness of the Swedish system, as explained above, conclusions should be transferred with caution.

8 Conclusion

By analyzing panel data on grade inflation in Swedish upper secondary school, we investigate the effects of a 2011 reform that introduced restructured curricula and a different grading scale. We find that the reform has increased grade inflation overall, but that the increase has been greater for public schools in relation to their independent counterparts. On an aggregate level, inflation has increased by 6.5–7%, while we find a significant increase of about 10% in Mathematics and roughly 12% in Swedish. As for the relative change between independent and public schools, there has been an aggregate decrease by 2–3%—inflation has increased for both school types, but more in public than in independent. When analyzing the subjects individually, the differences have decreased by 2.8% in Mathematics and 3.6% in Swedish. For English courses, we find no significant change in absolute grade inflation or relative difference between school types.

Considering these findings and the purpose of this paper, has the reform impacted grade inflation in the right direction? Even though levels have risen substantially—and differently between subjects—since the implementation began, we want to stress that this does not automatically reject the reform as inferior. The shift does not necessarily stem from lower grading standards, but could be a result of how the new system is constructed. The contribution of each factor is beyond the scope of our research. Inflated grades are not adverse when all schools practice it uniformly. The prime concern is instead if grades are inflated systematically and unjustly, since this both hurts quality in education and amplifies information asymmetries between students. On an aggregate level, the reform appears to have reduced inequalities between independent and public schools. On the one hand, one could infer that SNAE has achieved its goal of alleviating differences. On the other hand, when analyzing the subjects separately, another story unfolds. Independent schools still inflate more than public in English and Mathematics, and the difference has only become smaller in Mathematics with 90% significance. Conversely, in Swedish courses, independent schools inflated less and this gap has widened significantly. All in all, the aggregate reduction in inequality seems to some extent be attributed to increased inequality in Swedish courses. In terms of effectiveness, the conclusion depends on the scope of fairness. Contemplating the aggregate results, the convergence translates into students' GPA from independent and public schools being more equivalent in reflecting achievement—a sign of effectiveness. However, when we take into account the separate subjects, we can only conclude effectiveness in Mathematics, while we can show no and even negative effects in English and Swedish respectively. The question of effectiveness becomes almost philosophical, as to whether the constituents or the whole is of highest importance. The final verdict whether the reform has succeeded becomes rather ambiguous. One must also keep in mind the limitations, which imply that we should be careful when drawing absolute conclusions.

Furthermore, our findings and analysis give rise to some practical implications for policy. Firstly, we identify increased central control as a useful tool to reduce unjust grading practices if implemented correctly. In the Swedish criterion-based grade system, where individual schools bear the entire responsibility for grading, both students and schools have an incentive to favor more lenient grading. Thereby, central control is crucial in preventing local discrepancies. Secondly, we suggest clearer guidelines to the use of national tests in final grading. Though student assessment should be nuanced, tests provide a useful indication of student achievement, and the data still reveals highly varying interpretations between schools.

Future research could, first of all, investigate similar questions asked in this paper over longer periods of time, in order to evaluate the long-term effects of the reform. E.g. is there a transition period that students and teachers have to pass through? Secondly, a closer examination of the different subjects could help understand the underlying differences that we observe but are unable to fully explain. Thirdly, just as previous research points out, the evident skew toward grade inflation in schools with high female percentages still remains unexplained and should be investigated.

9 Summary

The purpose of this paper is to describe how the reform of 2011 has impacted grade inflation in Swedish upper secondary school and evaluate its ability to alleviate differences in grading practices between independent and private schools.

By analyzing data on the relationship between final grades and national test performance, we summarize our contribution to this field of research as follows:

1. The absolute levels of grade inflation, as we define it, have increased with the reform. This trend is consistently significant in Mathematics and Swedish, but not in English. We cannot determine if this increase is due to lowered grading standards. We also reveal patterns in why grade inflation differ between different courses.

2. We find that, aggregately, independent schools' tendency to inflate grades more than public schools has decreased with the reform. We attribute this betterment to the reform.

3. The shift is not uniform between subjects, and while we find significant evidence for a shrinking gap in Mathematics, public schools' higher grade inflation in Swedish courses increases after the reform.

In closing, this paper identifies an evident increase in grade inflation, but more ambiguous conclusion in terms of fairness. We choose to leave it up to parents, students, teachers, and politicians to determine what is really fair. We prove, however, that the Swedish National Agency for Education's efforts have made *a difference* with the reform—for better and worse.

10 References

Bedard, K. 2001, "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy*, vol. 109, no. 4, pp. 749-775.

Bettinger, E. 2011, "Educational Vouchers in International Contexts" in *Handbook of the Economics of Education*, eds. E.A. Hanushek, S. Machin & L. Woessmann, 4th edn, North-Holland, Amsterdam, pp. 551-572.

Björklund, A., Fredriksson, P., Gustafsson, J. & Öckert, B. 2010:13, *Den svenska utbildningspolitikens arbetsmarknadseffekter: vad säger forskningen?* [The Labour Market Effects of the Swedish Education Policy: What Does the Research Say?], Institute for Evaluation of Labour Market and Education Policy, Uppsala.

Black, P. & William, D. 2005, "Lessons from Around the World: How Policies, Politics and Cultures Constrain and Afford Assessment Practices," *Curriculum Journal*, vol. 16, no. 2, pp. 249-261.

Bonesrønning, H. 2004, "Do the Teachers' Grading Practices Affect Student Achievement?," *Education Economics*, vol. 12, no. 2, pp. 151-167.

Böhlmark, A., Holmlund, H. & Lindahl, M. 2016, "Parental Choice, Neighbourhood Segregation or Cream Skimming? An Analysis of School Segregation After a Generalized Choice Reform," *Journal of Population Economics*, vol. 29, no.2, pp. 1-36.

Bunar, N. 2010, "Choosing for Quality or Inequality: Current Perspectives on the Implementation of School Choice Policy in Sweden," *Journal of Education Policy*, vol. 25, no. 1, pp. 1-18.

Figlio, D.N. & Lucas, M.E. 2004, "Do High Grading Standards Affect Student Performance?," *Journal of Public Economics*, vol. 88, no. 9-10, pp. 1815-1834.

Jacob, B. & Levitt, S. 2003, "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, vol. 118, no. 3, pp. 843-877.

Kanagaretnam, K., Mathieu, R. & Thevaranjan, A. 2003, "An Economic Analysis of the Use of Student Evaluations: Implications for Universities," *Managerial and Decision Economics*, vol. 24, no. 1, pp. 1-13.

Langbein, L. 2008, "Management by Results: Student Evaluation of Faculty Teaching and the Mis-measurement of Performance," *Economics of Education Review*, vol. 27, no. 4, pp. 417-428.

Nata, G., Pereira, M.J. & Neves, T. 2014, "Unfairness in Access to Higher Education: A 11 year Comparison of Grade Inflation by Private and Public Secondary Schools in Portugal," *Higher Education*, vol. 68, no. 6, pp. 851-874.

Sabot, R. & Wakeman-Linn, J. 1991, "Grade Inflation and Course Choice," *Journal of Economic Perspectives*, vol. 5, no. 1, pp. 159-170.

SOU 1992:86, *Ett nytt betygssystem* [A New Grading System], Ministry of Education

and Science, Stockholm.

Spence, M. 1973, "Job Market Signaling," *Quarterly Journal of Economics*, vol. 87, no. 3, pp. 355-374.

Swedish National Agency for Education 2009, *Likvärdig betygssättning i gymnasieskolan? En analys av sambandet mellan nationella prov och kursbetyg* [Fair Grading in the Upper Secondary School? An analysis on the relationship between national tests and course grades], Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2011, *Gymnasieskola 2011 [Upper Secondary School 2011]*, Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2012, *Redovisning av uppdrag om avvikelser mellan prov-resultat och kursbetyg i gymnasieskolan* [Commissioned Account of Differences between Test Scores and Course Grades in Upper Secondary School], Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2013, *Redovisning av uppdrag om avvikelser mellan prov-resultat och kursbetyg i gymnasieskolan* [Commissioned Account of Differences between Test Scores and Course Grades in Upper Secondary School], Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2014a, *Lärarnas uppfattningar om 2011 års gymnasiereform – ett år efter införandet* [The Sentiment of the Teachers About the Upper Secondary School Reform of 2011], Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2014b, *Redovisning av uppdrag om avvikelser mellan prov-resultat och kursbetyg i gymnasieskolan* [Commissioned Account of Differences between Test Scores and Course Grades in Upper Secondary School], Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2015, *Redovisning av uppdrag om avvikelser mellan prov-resultat och kursbetyg i gymnasieskolan* [Commissioned Account of Differences between Test Scores and Course Grades in Upper Secondary School], Swedish National Agency for Education, Stockholm.

Swedish National Agency for Education 2016a, *SIRIS Database* [Homepage of Swedish National Agency for Education], [Online]. Available: `http://siris.skolverket.se/siris` [Extracted March 1, 2016].

Swedish National Agency for Education 2016b, *Utvärdering av den nya betygsskalan samt kunskapskravens utformning* [Evaluation of the New Grading Scale and the Criteria Layout], Swedish National Agency for Education, Stockholm.

Swedish Schools Inspectorate 2013, *Olikheterna är för stora* [The Differences Are Too Large], Swedish Schools Inspectorate, Stockholm.

Söderström, M. & Uusitalo, R. 2010, "School choice and segregation: Evidence from an admission reform," *Scandinavian Journal of Economics*, vol. 112, no. 1, pp. 55-76.

Timbro 2013, *Vinstdrivande skolor* [For-profit Schools], Timbro, Stockholm.

Vlachos, J. 2010, *Betygens värde: En analys av hur konkurrens påverkar betygsättningen vid svenska skolor* [The Value of the Grades: An Analysis of How Competition Affects Grading in Swedish Schools], Swedish Competition Authority, Stockholm.

Waldow, F. 2014, "Conception of Justice in the Examination Systems of England, Germany, and Sweden: A Look at Safeguards of Fair Procedure and Possibilities of Appeal," *Comparative Education Review*, vol. 58, no. 2, pp. 322-343.

Walsh, P. 2010, "Does Competition among Schools Encourage Grade Inflation?," *Journal of School Choice*, vol. 4, no. 2, pp. 149-173.

Wikström, C. 2005a, *Criterion-referenced Measurement for Educational Evaluation and Selection*, Department of Educational Measurement at Umeå University.

Wikström, C. 2005b, "Grade Stability in a Criterion-referenced Grading System: The Swedish Example," *Assessment in education: Principles, Policy & Practice*, vol. 12, no. 2, pp. 125-144.

Wikström, C. & Wikström, M. 2005, "Grade Inflation and School Competition: An Empirical Analysis Based on the Swedish Upper Secondary Schools," *Economics of Education Review*, vol. 24, no. 3, pp. 309-322.

William, C., Li, H. & Wing, S. 2007, "A Signaling Theory of Grade Inflation," *International Economic Review*, vol. 48, no. 3, pp. 1065-1090.

Wilson, B.P. 1998, *The Phenomenon of Grade Inflation in Higher Education*, Condensed speech edn, Virginia Professional Educators, Fredericksburg, Va.

Wooldridge, Jeffrey M. 2010, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Mass: MIT Press.

Yang, H. & Yip, C.S 2003. An Economic Theory of Grade Inflation. *Unpublished manuscript, Department of Economics, University of Pennsylvania*, Philadelphia, Pa., Nov. 4, 2003.

Ziomek, R.L. & Svec, J.C. 1997, "High School Grades and Achievement: Evidence of Grade Inflation," *NASSP Bulletin*, vol. 81, no. 587, pp. 105-113.

Appendices

A Summary statistics

Table 6: Summary statistics

Variable	Mean	Std. Dev.	N
Grade inflation	0.174	0.223	20890
9^{th} grade GPA	217.613	30.795	20307
Female percentage	48.187	21.212	20396
Independent percentage	38	27.1	20773
Entering students municipality	2390.383	3341.462	20890
Schools per student	0.004	0.003	20882
Students in school	354.63	312.676	20396
Enter index	0.922	0.131	20868
Students per teacher	12.845	4.149	20204
Educated parents percentage	46.949	16.414	20396
Foreign background percentage	18.876	15.478	20396

B Correlation matrix

Table 7: Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Grade inflation	1.000															
(2) Year	0.046	1.000														
(3) 9^{th} grade GPA	-0.023	0.086	1.000													
(4) Mathematics	0.364	-0.000	0.000	1.000												
(5) Swedish	0.036	-0.000	0.000	-0.500	1.000											
(6) Students per teacher	0.042	-0.007	0.304	0.000	-0.000	1.000										
(7) Female percentage	0.072	-0.008	0.250	0.000	0.000	0.137	1.000									
(8) Students in school	-0.010	-0.188	0.202	-0.000	-0.000	0.164	0.023	1.000								
(9) Educated parents percentage	-0.037	0.066	0.748	-0.000	0.000	0.353	0.160	0.207	1.000							
(10) Foreign background percentage	0.050	0.108	-0.041	0.000	0.000	0.165	0.055	0.098	-0.126	1.000						
(11) Independent percentage	0.004	-0.185	0.083	0.000	0.000	0.292	0.090	0.126	0.203	0.128	1.000					
(12) Entering students municipality	-0.021	-0.017	0.198	0.000	0.000	0.349	0.083	0.054	0.277	0.322	0.400	1.000				
(13) Enter index	0.026	-0.273	-0.008	-0.000	-0.000	0.037	0.022	0.133	-0.018	0.011	0.154	0.087	1.000			
(14) Grade level	0.174	-0.047	0.121	-0.000	0.000	0.066	0.008	0.120	0.140	0.025	-0.005	0.007	0.010	1.000		
(15) Independent	0.030	-0.084	-0.078	-0.000	-0.000	0.224	0.036	-0.301	0.053	-0.045	0.543	0.268	0.080	-0.043	1.000	
(16) Reform	-0.012	0.549	0.013	-0.000	0.000	-0.001	0.005	-0.169	0.022	0.060	-0.109	-0.009	-0.181	-0.518	-0.024	1.000

C Proportions of deviations across subjects

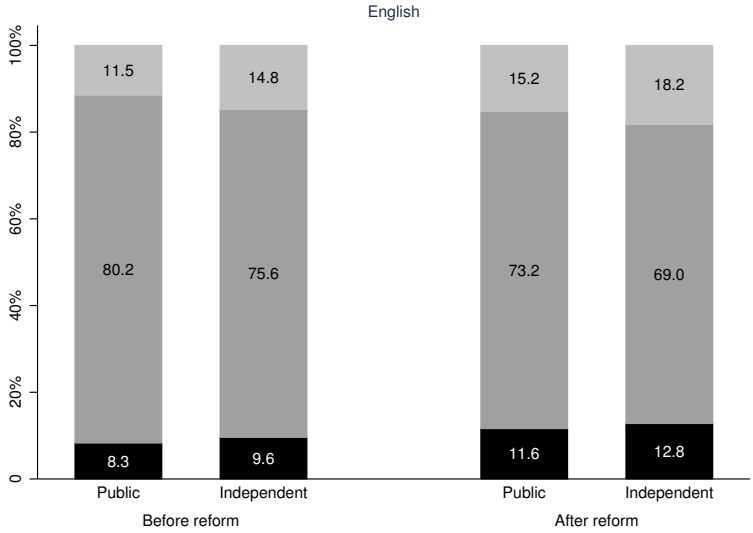

Figure 7: Average proportions of students receiving lower (black), equal (dark gray), and higher (light gray) final course grades in English relative national test, before and after the reform. Averages are grade level-weighted. Independent schools inflate grades more before the reform. Inflation increases little for both, but to a smaller extent in independent schools.

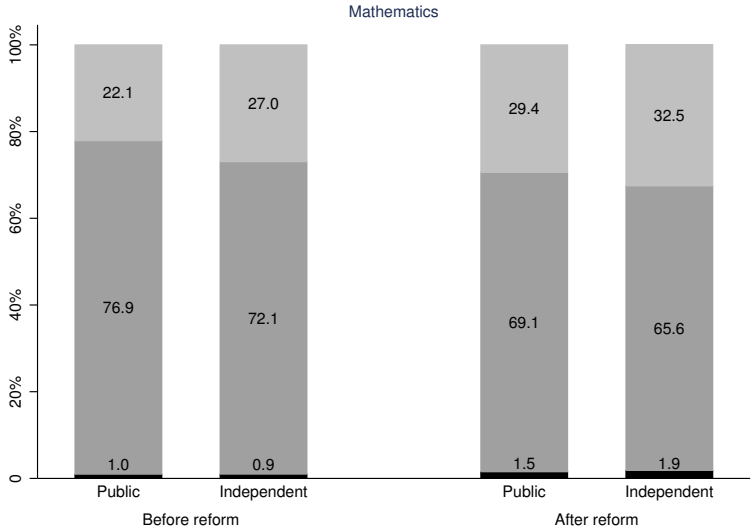

Figure 8: Average proportions of students receiving lower (black), equal (dark gray), and higher (light gray) final course grades in Mathematics relative national test, before and after the reform. Averages are grade level-weighted. Independent schools inflate grades more before the reform. Inflation increases for both, but to a smaller extent in independent schools.

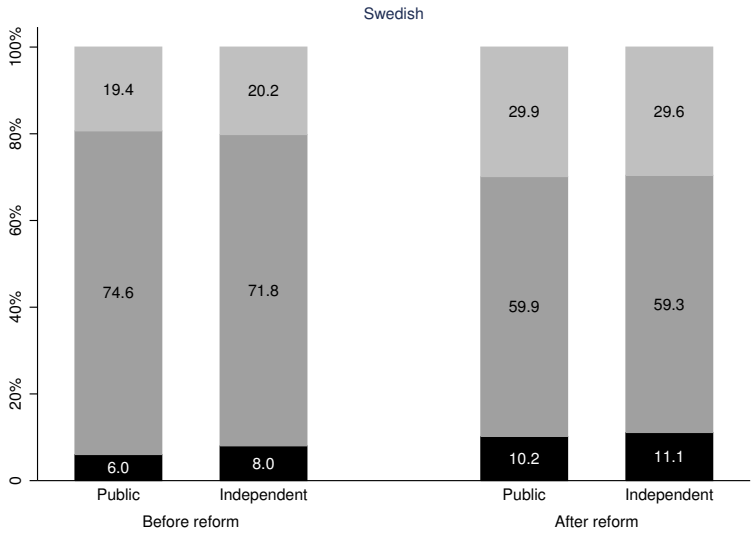

Figure 9: Average proportions of students receiving lower (black), equal (dark gray), and higher (light gray) final course grades in Swedish relative national test, before and after the reform. Averages are grade level-weighted. Public schools inflate grades more before the reform. Inflation increases for both, but to a smaller extent in independent schools.