Lund University

Master Thesis

A for effort: an empirical analysis of the effect of competition on grade inflation in Swedish schools

Author:
Hampus Sporre

Supervisor:
Alessandro Martinello

May 2016

Abstract:

I present a theoretical model for competitive interaction between schools and grade inflation. To test the model, I estimate the effect of competition on grade inflation in Swedish upper secondary schools using data from all schools in the years 2012-2015. To solve the endogeneity problem of competition, I use Instrumental Variable estimation with Fixed Effects. I use the sudden shutdown of 17 schools due to the bankruptcy of the John Bauer Group in 2013 as an instrument for competition. The IV results and the exploration of an adverse effect of grade inflation both support the theoretical model. The IV results show that there is a causal effect of competition on grade inflation.

Keywords: Grade inflation, competition, education, instrumental variables, fixed effects.

I want to thank Erik Mohlin for contribution to the model and Håkan J. Holm for valuable comments in the early stages of this project.

1 Introduction

Education has a central role in the welfare state of Sweden. Public funding and profits trigger both ideological debates and efficiency concerns. Declining performance in international comparisons despite firmly soaring grades has made grade inflation a common assertion. Claims that the high degree of competition in the schooling sector incentivises grade inflation have not been absent. If competition causes grade inflation, competition causes problems.

Research on competition in the schooling sector and its effect on educational quality is plentiful (e.g. Hsieh and Urquiola 2006, Sandström and Bergström 2005, Hoxby 1994). It is widely anticipated that competition increases educational quality. However, results are mixed not only because schooling is different in different countries, but also because educational quality is difficult to measure.

To use grades as a measure of educational quality is only valid if they signal true learning outcomes. If grades are inflated, they do not signal true learning outcome. If the examiner has an interest in the grade she sets, it is likely that she inflates it (Jacob and Levitt 2003). Björklund et al. (2010) shows that the grades of Swedish students have soared steadily since the mid 1990's in a manner that is unreasonable to reflect true learning development. Vlachos (2016) compares the distributions of grades in upper secondary school (high school) in 1998 and 2008. The distribution remarkably shifts upwards during these years. Also, the fraction of students that receives maximum grade average increases extensively. Altogether, these papers suggest that grade inflation has increased severely in Sweden.

Except for disqualifying grades as an appropriate measure of true educational quality, grade inflation causes several detrimental effects on the schooling sector and the labour market. First, sorting into career paths fails. In Sweden, grades are important for admission to higher education. Because grades are the main admission instrument to higher education, it is important that grades are reliable. If grades are inflated, the admission is not legitimate and the sorting into higher education fails. Students following the wrong career paths can be very costly to society (Kostal et al. 2016).

Second, the legitimacy of grades as an admission instrument declines. Grades are typically good predictors of future educational success. Björklund et al (2010, Chapter 10.3) shows that grades are superior to alternative admission instruments (e.g. SweSAT) in predicting learning outcomes in higher education. If grades are inflated and do not signal true learning achievement, then grades do not predict future educational success. If the legitimacy of grades declines, some other admission instrument has to replace grades. Discarding the superior predictor of educational success as admission instrument is an obvious deterioration for higher education.

Third, grade inflation causes bad matchings on the labour market. Grades are important matchmakers on the labour market. Because grades signal ability, employers frequently use grades to distinguish between applicants (Vlachos 2010, Wikström and Wikström 2005). If grades are inflated, they do not signal ability anymore. Therefore, grade inflation makes it more difficult for high ability applicants to stand out and prove their worth. If employers cannot discriminate between applicants, bad matchings on the labour market are to expect.

Finally, grade inflation is harmful for educational quality. The most important input to the education production is student effort (Bonesrönning 2004). Betts and Grogger (2003) finds that inflated grades lead to less student effort. If a student experiences that grades are easily attained, she learns that effort is not necessary. Therefore, grade inflation seriously damage educational quality and learning outcomes (Vlachos 2010).

There is a conflict of interest between the financier and the student. Because high grades ensure admission to higher education and signals ability to employers, students prefer high grades. Because upper secondary schooling is publicly funded in Sweden, the public wants fair grading so that sorting into higher education is legitimate (Figlio and Loeb 2010).

The provider can however respond to the students' interests for the school. Upper secondary schooling in Sweden is publicly funded, regardless of who provides it. Most municipalities provide public schooling. However, profit maximizing firms providing independent schools are common. Because upper secondary education is funded publicly, the choice of public or independent schooling is detached from costs for the student. To raise profits, the provider needs to attract students. Therefore, it is the interests of the students that incentivises the education provider. Because the incentives are such that inflated grades are profitable to the grader, competition between schools ought to spur grade inflation (Vlachos 2010, Wikström and Wikström 2005).

This paper tests if competition between upper secondary schools affects grade inflation. I use data on scores on national standardized tests and grades from all upper secondary schools in Sweden to measure grade inflation. I use data on enrollment to calculate market concentration as a measure of competition. To identify the causal effect of competition, I exploit the bankruptcy of the John Bauer Group in 2013. More specifically, I use Instrumental Variable (IV) estimation to tackle the problem of simultaneous effects. I use the shutdown of 17 schools due to the bankruptcy as an instrument for competition. I use fixed effects to control for time invariant omitted variables. The results show that competition triggers grade inflation.

I contribute to the literature on competition and grade inflation in several aspects. First, while most studies investigate elementary schools, I instead consider the market for upper secondary schooling. Second, I combine the most appropriate measure of competition with the most accurate measure of grade inflation. Third, I use data from a later time period, exploiting the new grade scale and more competition. Finally, I use a more careful empirical strategy, admitting several sources of the endogeneity of competition and counter them.

I also provide a theoretical model of the competitive interaction between schools. To further test the details of this model, I explore the negative effect of grade inflation on learning outcomes, suggested in the literature. I use this mechanism to identify severe grade inflators and investigate what characteristics these schools have. This is an additional analysis that aims to evaluate the theoretical model and investigate the suggested (by the literature) mechanism. This analysis coarsely confirms the predictions of the model and provides some additional understanding for the empirical results.

The remainder of this paper is structured as follows. Section 2 describes the Swedish system for upper secondary schooling and how the literature suggests that competition relates to grade inflation. Section 3 presents the theoretical model and its predictions for grade inflation. Section 4 describes the data, how I estimate the effect of competition on grade inflation and the results. Section 5 explores the negative effect of grade inflation on learning outcomes and analyses the characteristics of grade inflating schools. Section 6 concludes.

2 School choice and grade inflation

This section explains the Swedish system for upper secondary schooling and discusses grade inflation. The first sub-section briefly describes school choice and grading. The second subsection discusses the relationship between competition and grade inflation and how it is studied in previous research. The second sub-section also discusses the methods and results of studies on competition and grade inflation and what the possible improvements are.

2.1 The Swedish system

In Sweden, upper secondary schooling is free. The full cost is borne by the student's residential municipality. Upper secondary school consists of three years, tenth through twelfth grade. Most of the municipalities provide public schools but since a major reform in the early 1990's, independent schools have become increasingly common. Independent schools are privately provided but they are in general free of charge, from the student's perspective. Each student chooses which school to attend and the residential municipality pays the school. Because the payment per student (Skolpeng) is fixed, the number of enrolled students determines the income of the school. There must be special reasons and approval from the Swedish National Agency for Education (Skolverket) for a school to charge tuition. However, permission to charge tuition is very rarely granted. As the individual student can attend any upper secondary school at no cost, the majority attends. Because the payment is publicly funded and fixed per student, the public and private schools compete on equal conditions. Independent schools have therefore become common and in the academic year of 2014/2015, about 25 per cent of the students attended independent schools.

Academic merits from elementary school determine if the student can enroll in her preferred school. Students apply to schools through a centralized system where the student ranks her preferred schools. The GPA from elementary school then determines which students that are enrolled. If the education requires certain skills within aesthetics, additional admission tests are allowed. However, most upper secondary educations do not use additional tests and therefore the GPA from elementary is the main admission instrument.

The individual teacher grades her students with discretion. In upper secondary school, the subjects are divided into courses, which last for about one to two semesters depending on the size of the course. A student takes several courses simultaneously, usually about ten per year. The teacher sets the course grade at the end of the course. The grade shall reflect the student's educational achievement based on performance during the course. The grade scale is A-F, where grades A-E are different levels of pass, A being the highest. This scale was introduced in 2011 and the previous system only had three different pass grades. There are criteria for each grade level but it is the individual teacher that evaluates the student and decides which grade is appropriate. Because the grade is based on performance during the whole course, the teacher is very free in applying the criteria and setting the grade.

The discretion of the grading raises questions about the legitimacy of grades. Because teachers are autonomous in the complex task of grading students, they carry great responsibility. When grading her student, the teacher is in some sense also grading herself. Vlachos (2010) points out that having teachers grading themselves creates skewed incentives for the teachers and it requires some external control. If teachers have incentives to set high grades and external control is low, grades are artificially inflated and the legitimacy of grades declines (Vlachos 2010).

Standardized tests do in some subjects help the teacher to assess the students. There are national standardized tests in the courses in Mathematics, Swedish and English. The test is an additional assessment device but also an instrument for comparison between schools. Experienced university staff with good knowledge of how to examine these courses constructs the tests. Therefore, the tests signal the educational achievement of students in a comparable manner that helps the teacher in assessing. However, the teacher is only supposed to use the test score as an additional assessment device when grading the student. The score does not determine the grade. It is also the teacher that marks the national tests and only a small sample is collected for external review. Because the test score does not determine the grade, the teacher is free to choose how much to rely on them.

The grades are important for the student's future. The course grades are combined in a measure that determines admission to higher education. The admission also has a quota for the Swedish Scholastic Aptitude Test (SweSAT), which is a test that can be taken by anyone and measures appropriate abilities that are needed for higher education. Except for this quota (1/3), admission is based on grades from upper secondary school.

2.2 The role of competition

The Swedish market for upper secondary schooling welcomes competition. Because the choice of school is nearly unrestricted and detached from costs, schools compete on equal terms. Therefore, the most attractive schools receive the most applicants. Because enrollment determines funding, the most attractive schools raise the most revenues.

Competition on the schooling market is problematic because of asymmetric information. It is not necessarily true educational quality that makes a school attractive and therefore profitable. Studies on what inputs create educational quality reach many different conclusions (e.g. Bonesrönning 2003). At the choice of school, the students have no experience and there is practically no room for trial and error learning. Because there is no objective measure of educational quality, the students must themselves evaluate the alternatives and base their decision on observables. When observables are not the true quality, and the school can manipulate the observables, the problem of asymmetric information rises. Akerlof (1970) shows that when consumers cannot observe true quality, the sellers of bad quality claim to be of high quality and drive high quality sellers out of the market. In the schooling sector, this translates into students having no possibility to distinguish between high and low quality schools. Even after graduation, students lack comparison and do not know if they attended a high quality education. Because educational quality is costly, the education provider profits on reducing quality but claiming to provide high quality. Hence, the informational advantage of the education provider is likely to have a negative effect on education quality (Vlachos 2010).

Students use the schools' grade levels to decide which school to attend. When deciding upon schools, students face many attributes of quality. Vlachos (2010) and Böhlmark and Lindahl (2008) suggest that the grade level is the main attribute available to students in evaluating educational quality. High grades signal good learning outcomes. Moreover, students need high grades for admissibility to higher education. Because grades are the main admission instrument to higher education, students value high grades even if they are inflated. Inflated grades actually make students eligible to higher education to lower costs in terms of effort. If students desire admissibility to higher education, high grades are attractive regardless of true learning outcome.

School choice and competition encourage grade inflation. Because enrollment determines funding, schools have the incentive to attract many students. As high grades are attractive to students, schools want to set high grades. The school can secure a high grade level in two ways. The first is to improve the educational quality, which is costly. The second is to inflate grades. The discretion of grading in Swedish schools makes grade inflation less costly, as the risk of detection is low. Hence, Swedish schools are anticipated to inflate grades (Vlachos 2010, Wikström and Wikström 2005, Böhlmark and Lindahl 2008).

The presence of grade inflation is a widespread problem across countries and it is not easily prevented. Costley (2014) shows that grade inflation is common practice even when external control is more extensive than in Sweden. Jacob and Levitt (2003) identifies straightforward cheating as teachers correct their students answer sheets before sending them for external grading. Therefore, it is likely that grades are inflated as long as there are incentives.

The Swedish system is a recipe for grade inflation. The discretion of grading and lack of external control reduces the risk of detection (Vlachos 2010, Wikström and Wikström 2005). The competition between schools and the problem with asymmetric information creates incentives for grade inflation. The Swedish Competition Authority acknowledges the potential effect from competition on grade inflation. The Swedish Competition Authority requested a study on the effect of competition on grade inflation. The study is Vlachos (2010), and it considers competition between elementary schools.

Measuring grade inflation is a complex task. There are several measures of grade inflation in the literature, but only one of them is compelling. First, comparing grades from different levels of schooling is not appropriate. Vlachos (2010) compares grades in upper secondary school with the grades from elementary school. The argument for comparison of grades from different levels is that grade inflation induces students to put less effort in their education (Bonesrönning 2004). As students put less effort, their grades should decline. There are two main problems with the comparison of grades from different levels. The first problem is that if grades are inflated, they do not necessarily indicate student effort. If grades are inflated differently in the different levels, such a comparison is misleading. An additional problem is that it assumes that grade inflation is the main determinant of the change in student performance. If something else affects the teenager's educational development, the precision of this measure declines.

Second, comparison between grades and the SweSAT score is not a desirable measure either. Wikström and Wikström (2005) measures grade inflation in upper secondary school by comparing each student's average grade with her score on the SweSAT. The SweSAT score is certainly comparable between students but it does not measure the same abilities that grades do. There are three main reasons why comparing a student's grades with her SweSAT score is problematic. First, scoring very high on the SweSAT early in upper secondary school assures the student that she will be admitted to higher education with her test score. As grades are no longer important to her, she is less likely to put effort in her remaining upper secondary schooling. If she does not put effort in her studies in response to her high SweSAT score, the comparison between her grades and SweSAT score is misleading (Vlachos 2010). The second problem is that because the SweSAT is optional, there is strong social selection into the sample. The third problem is that the test score tends to increase with the number of times the test is taken (Björklund et al. 2010). Because of these problems, comparing students' grades with their SweSAT score is not an appropriate measure of grade inflation.

Finally, the relationship between grades and national standardized test scores can be used to measure grade inflation. Vlachos (2010) and Böhlmark and Lindahl (2008) use the relationship between grades and scores on the national standardized tests to measure grade inflation in elementary school. This measure assumes that test scores signal true learning achievement. Because the tests cannot measure every assesment criterion on the course, the score does not determine the grade. While some students have a bad day when taking the test, some are lucky. Therefore, the test result of a student is not necessarily the grade she deserves. However, on average these differences should cancel out, as there is no reason for general under- or over performance on the test.

A problem with the difference between the grade and test score as a measure of grade inflation is that it is the same teacher that sets the grade and marks the national test. Because teachers are likely to inflate grades, they can as well inflate the test scores (Figlio and Loeb 2010, Jacob and Levitt 2003). Vlachos (2010) argues that grades are much easier to inflate, and therefore the measure is still valid but underestimates the true grade inflation.

Because of high precision and only minor problems (underestimation), I use the relationship between grades and national test scores to measure grade inflation.

Measuring competition is also complicated. To measure competition, I must first define the market. Because few students attend schooling in another municipality than the resident, all Swedish studies mentioned in this paper define the market as the municipality (e.g. Böhlmark and Lindahl 2008). Tiebout sorting could cause competition to reach across municipalities as well. Black and Machin (2011) shows that school standards affect house valuations in many countries, but Sweden is not included in the study. While Vlachos (2010) warns for inter-municipality competition, Ahlin and Johansson (2000) shows no sign of Tiebout sorting in Sweden. I follow the literature and define the market as the municipality.

How to quantify the degree of competition in a municipality is also a critical choice. The most common measure of competition in the literature is the share of students that attend an independent school (e.g. Vlachos 2010, Sandström and Bergström 2005, Ahlin 2003). If independent schools gaining larger market shares are the main drivers of competition, this measure is valid. However, market structures and competition can be very different across municipalities even if they have the same share of students in independent schools. Consider two different municipalities. One of them has two schools of equal size, one public and one independent. Hence, the share of students in independent schools is 0.5. The other municipality has ten different schools of different size, but the share of students in independent schools is also 0.5. The municipality with ten small schools is likely to be more competitive than the municipality with only two. If the municipality with ten small schools is more competitive than the municipality with only two schools, the share of students in independent schools fails to measure competition.

A proper measure of competition takes competition between public schools into account as well. Wikström and Wikström (2005) tries to incorporate competition between public schools in its measure of competition. This measure is a classification of each municipality depending on if there are independent schools and potential competition between public schools. As long as municipalities only have a few schools and the main difference is if there are multiple public schools or an independent school, this measure is useful. However, if there are many schools in each municipality, a simple classification is also a poor measure of competition.

Standard measures of competition (that do not use price elasticities and markups) account for the competitors' market shares and relative sizes. As Sandström and Bergström (2005) suggests, the Herfindahl-Hirschmann Index (HHI) is a "more natural way, to measure competition between municipal schools". Because the HHI is the sum of squared market shares, it accounts for the schools' relative sizes and captures the market concentration in a comparable measure. Walsh (2010) also recognizes the suitability of the HHI when studying competition on the schooling market in the US. I follow these recommendations and use the HHI to measure competition in this paper.

Results from previous studies must be interpreted with care. Wikström and Wikström (2005) is the only study of competition and grade inflation in the Swedish upper secondary school that I am aware of. The measurements Wikström and Wikström (2005) uses are as mentioned problematic. The method is regression analysis with covariates, which reveals a causal effect only if the competition variable, conditional on covariates, is exogenous. Because I find it possible that grade inflation affects competition, I will relax this assumption. For example, if grades are not inflated, a potential entrant can easily enter the market and gain market share by inflating grades. If the degree of grade inflation affects the attractiveness to an entrant, grade inflation has an effect on competition. This effect of grade inflation on competition creates a problem of simultaneity. If there is a problem of simultaneity, the control strategy of Wikström and Wikström (2005) is flawed.

The literature suggests measures like the HHI for measuring competition between schools. Wikström and Wikström (2005) suggests a market concentration based measure of competition, which the HHI is. Because competition between public schools possibly also triggers grade inflation, Vlachos (2010) requests a measure that incorporates competition between public schools. The HHI accounts for competition between all schools.

Wikström and Wikström (2005) also speculates that teachers in larger schools have less incentive to inflate grades because they constitute a smaller fraction of the school's total average grade level. Wikström and Wikström (2005) also supposes that larger schools might have internal mechanisms to prevent grade inflation. I explore this suggestion in Section 5.

3 A theoretical model

It is desirable to have a theoretical model that describes schools' behaviour in a competitive environment. However, I find no such model in the literature. I hereby present a stylized model for the grading practice when schools interact competitively.

To define the profit function of the school, I define an expression that tells how the school receives funding, which is followed by different cost components associated to the inputs. The model is the following:

$$\pi_i = \frac{e^{\theta(k_i+g_i)}}{\sum_j^n e^{\theta(k_j+g_j)}} M - \alpha\frac{k_i^2}{2} - \phi\frac{g_i^2}{2} - \gamma\sum_j^n (s_j)^2 \frac{g_i^2}{2} \tag{1}$$

The first expression assigns a market share to school i depending on θ, k_i and g_i, in relation to k and g of all j schools in the municipality. θ is the degree to which grades influence the choice over schools, k_i is the true knowledge level of the students in school i and g_i is the level of grade inflation in school i. The sum $k_i + g_i$ is the observed grade level that students base their choice on. If θ is infinitely large, i.e. grade level is the only characteristic that students care about, the school with the highest grade level captures the whole market (M). If the grade level is not important at all ($\theta = 0$), the market shares are uniformly distributed across the schools. Because the payment to the education provider is fixed per student, this expression is a reasonable way to model the revenue incentive for the individual school.

The second term is a cost function of educational quality, i.e. the cost of true knowledge provision. I assume that it is more costly to improve upon an already high level of knowledge than it is upon a low level. Therefore, this function is exponential, i.e. the marginal cost of knowledge is increasing. The parameter α determines how costly educational quality is. If it is low, the provision of knowledge is not as costly as if the parameter is high.

The third term describes a cost of grade inflation that depends on ϕ, a parameter that describes the morale and professional conduct of the teachers on the school. If this parameter is high, grade inflation infers a higher cost in terms of immoral conduct and bad conscience. The marginal cost is increasing, assuming that a little grade inflation is acceptable but when it increases, the conscience of the teacher makes itself heard with more distinction.

Lastly, the model includes a cost of generosity that depends on the degree of competition. This cost is an (perceived) expected cost, which consists of a probability of detection, and the cost of punishment. The parameter γ defines the likelihood of detection and the punishment cost, and it decreases as the competition increases (HHI decreases). As the market concentration decreases, i.e. competition increases, it threatens the employment of the teacher (and the revenues of the school). The risk of the school getting punished becomes relatively less frightening to the teacher as the risk of unemployment rises. Also, there is a hide-in-the-crowd effect. If the competition increases, the individual teacher can "hide in the crowd", feeling less detectable in her grading. Hence, the perceived expected punishment should decrease with the degree of competition.

Maximizing the function with respect to g_i and k_i, respectively, gives the first order conditions:

$$\frac{\partial \pi_i}{\partial g_i} = \frac{M\theta e^{\theta(k_i+g_i)}(\sum_j^n e^{\theta(k_j+g_j)} - e^{\theta(k_i+g_i)})}{(\sum_j^n e^{\theta(k_j+g_j)})^2} - \phi g_i - \gamma \sum_j^n (s_j)^2 g_i \quad = 0 \tag{2}$$

$$\frac{\partial \pi_i}{\partial k_i} = \frac{M\theta e^{\theta(k_i+g_i)}(\sum_j^n e^{\theta(k_j+g_j)} - e^{\theta(k_i+g_i)})}{(\sum_j^n e^{\theta(k_j+g_j)})^2} - \alpha k_i \quad = 0 \tag{3}$$

Because it is the observed grade level, i.e. the sum of true knowledge and grade inflation, that determines the assignment of market shares, the first term is the same in both equations. Equalizing the cost terms from the first order conditions gives

$$\alpha k_i = g_i(\phi + \gamma \sum_j^n (s_j)^2) \tag{4}$$

I then substitute the right hand side in (4) for αk_i in (3). I consider only symmetric equilibria for simplicity reason, so that $i = j$ for all j, and solve for g^*. In the symmetric equilibria, all schools attain the same market share so the HHI translates into $n(\frac{1}{n})^2$, providing the following expression for the equilibrium level of grade inflation:

$$g^* = \frac{\theta M}{(\phi n + \gamma)}(1 - \frac{1}{n}) \tag{5}$$

The equilibrium level of true educational quality is given by

$$k^* = \frac{\theta M}{\alpha n}(1 - \frac{1}{n}) \tag{6}$$

To analyse how grade inflation is determined in the symmetric equilibrium of this simple model, I use (5). At first, it tells us that the more the students favour grades when they choose school (higher θ), the more are grades inflated. Because market shares become more valuable as the market size increases, there is also a positive effect from market size. Teachers' morale and expected punishment both reduce grade inflation. Note that the fraction $\frac{1}{n}$ is actually the measure of market concentration (HHI) in the symmetric equilibria:

$$\frac{1}{n} = n(\frac{1}{n})^2 \tag{7}$$

This means that grade inflation decreases with market concentration, i.e. it increases with the degree of competition, but only up to some point. Because n also appears in the denominator with ϕ, increasing the number of schools after a certain number actually reduces grade inflation. At which n this point is depends on the relative size of ϕ and γ. The smaller ϕ relative to γ, the higher n will bound the positive effect from competition on grade inflation. Since these parameters are hard to quantify, no such bound is suggested here. Because the incentive from unemployment risk is probably much stronger than the incentive of bad conscience, I assume that ϕ is relatively close to zero. This suggests that the alarm for competition and grade inflation is real. The results in Section 5 support these assumptions and predictions.

I use (6) to analyse how true educational quality is determined in the symmetric equilibrium. As with grade inflation, the importance of grades in the choice of school has a positive effect on the educational quality. The cost of quality naturally reduces it and market size increases it, similar to grade inflation. However, competition reduces educational quality, because marginal revenue from true knowledge decreases with the number of schools.

The main conclusion is that competition increases grade inflation. This theoretical analysis is very simple and it is presented only to give a theoretical framework. This model suggests that the degree of competition increases grade inflation. The following section shows empirical testing of this hypothesis.

4 Empirical analysis

This section presents the empirical framework and the estimated effects of competition on grade inflation. The first sub-section describes the data. The second sub-section is a step-by-step clarification of the procedure for estimation. The third sub-section presents the results.

4.1 Data

I use data from the Swedish National Agency for Education's (Skolverket) online information system – SIRIS. The data consist of reports from national tests, grades, enrollment and teacher information during the academic years of 2012/2013, 2013/2014 and 2014/2015. I refer to an academic year by its ending year, i.e. I refer to the academic year 2013/2014 as year 2014. Each observation is a test taken at a school that year, e.g. Mathematics 1C (mathematics for students in natural sciences at first level) at Katedralskolan in Lund at year 2015. In the sample there are 19 543 tests reported from 873 schools in 230 municipalities (all upper secondary schools in Sweden). By using data from these years, I exploit the new grade scale that was introduced in 2011. More grade levels make each student more likely to receive a different grade than test score. Because the new scale has five levels and the previous scale has only three, the new scale is more accurate in identifying grade inflation.

To measure grade inflation I use the net share of students in each test group that receive a higher grade than their score on the national test. I use this measure because the more students receiving higher grades than test scores, the more grade inflation there is. This assumes that the test score is the true achievement level of a student and that this should determine the true and fair grade. Because a student can have a bad day and grades are based on more criteria than can be measured on a test, there are legit individual deviations. However, because observations consist of groups of students, deviations should cancel out. Some students deserve higher grades than their test scores and some students deserve lower, there is no reason for general under performance on the test. Hence, if the majority of students in a group receive a higher grade than test score, the teacher is inflating grades. Therefore, this measure is a good proxy for grade inflation. Because some students receive lower grades than test scores, I calculate the net share. For example, if 30 per cent of the students taking a test receive higher grades and 20 per cent receive lower grades, the net share receiving higher grades is 10 per cent. The scale of this variable ranges from -1 to 1, so in our example where the net share that receives a higher grade is 10 per cent, the variable takes the value 0.1.

As mentioned earlier, test scores can be manipulated as well, as the generous grader can be generous also when marking the national tests. However, inflating test scores is more detectable than inflating grades because test scores are easier to review and a small random sample of the tests is actually collected and reviewed centrally at the National Agency of Education. Therefore, this measure is appropriate but it underestimates the true grade inflation.

Table 1 presents descriptive statistics on the level of observation. The bottom row shows that there are on average 52 students in each test group. I also present the average score (scaling from 0 to 20, minimum for pass is 10) and the share of the tests that are taken on a theoretical programme. The average net share of higher grades is 0.186, suggesting that grades are in general inflated. The columns for Higher and Lower present the means of the shares of students that receives higher and lower grades, respectively (i.e. the decomposed net share of higher grades). If less than 10 students in a school take a test and get graded in that course, this test is not included in the sample.

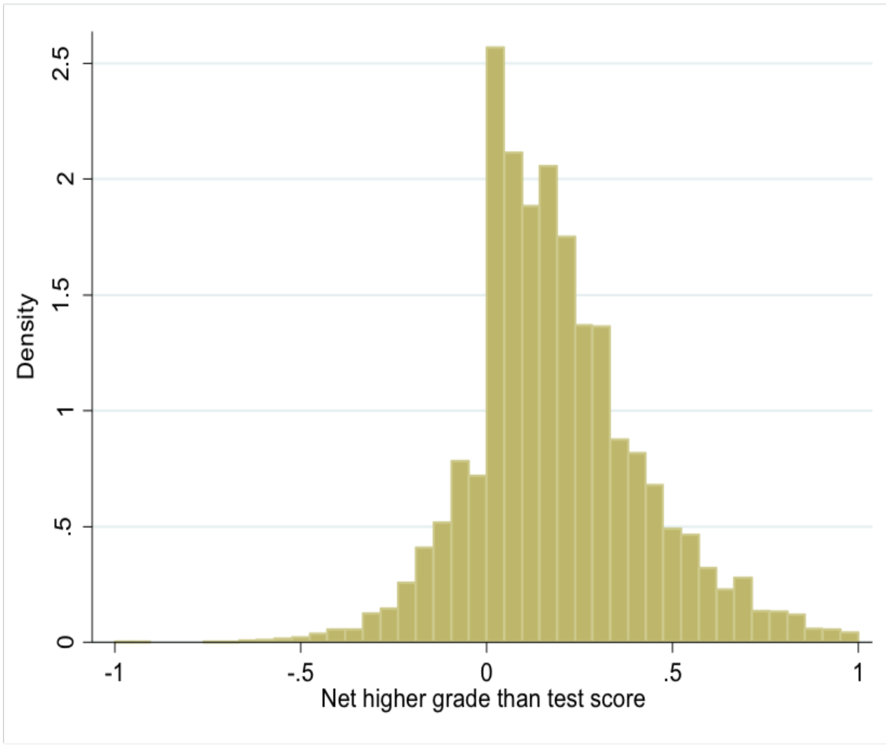
There are noticeable differences in the net share of higher grades across the subjects. This is probably because the test score is easier to inflate in some subjects than others (see Section 5 for extensive discussion).

Table 1: Test level descriptive statistics

	Means					
	Net higher	Higher	Lower	Students	Score	Theoretical
English	0.040	0.161	0.121	49.215	12.201	0.535
Math	0.268	0.284	0.016	55.006	12.421	0.625
Swedish	0.209	0.303	0.094	51.800	12.106	0.538
All tests	0.186	0.254	0.068	52.457	12.271	0.575

Figure 1 shows the distribution of the net share of students that receive higher grades than test scores. Because luck and bad luck on the test should be equally common, the share of higher grades than test score should be as large as the share of lower grades. Therefore, in absence of grade inflation, this distribution should be normal and centered on zero. The variance should be low if individual deviations cancel out. The observed distribution suggests three appearances. First, that the mean is above zero implies grade inflation. Second, the positive skewness implies that test scores pull up grades much more than they prevent them from soaring. Third, the discontinuity at zero suggests that something strongly prevents net shares from falling below zero. It is likely the convenience for the teacher to report in a zero net share that lifts up grades in schools where the net share should be negative. However, this “zero net effect” does not seem to restrain the teachers in classes where the net share lands slightly above zero. The national test is supposed to be an additional assessment tool for the teacher. However, Figure 1 suggests that test scores are mainly used when scores are higher than true achievement, but not so much when scores are low. To explore if competition affects the share of higher grades and the share of lower grades differently, I also estimate the effect on these shares separately.

Figure 1: Distribution of the net share that receives higher grade than test score



To measure competition, I follow the suggestion of Wikström and Wikström (2005) and consider market concentration, using the HHI. The HHI is a measure of market concentration and is therefore inversely related to competition (high HHI means low competition). In accordance to the literature, I define the market as the municipality, despite some students attending schooling in a different municipality than the resident (e.g. Vlachos 2010, Böhlmark and Lindahl 2008, Wikström and Wikström 2005).

Table 2 shows descriptive statistics on the level of municipality. I calculate each school's market share based on how many students it enrolls in relation to the total number of students in the municipality. The HHI then sums up the squared market shares in each municipality. The average HHI is 0.74 ($HHI \in (0, 1]$) but we also observe that almost half of the municipalities are monopolies ($HHI = 1$). Considering only non-monopoly municipalities, the market concentration reduces to 0.49 on average. As an example of a municipality with HHI close to the average of non-monopolies, consider Varberg, with 5-6 schools depending on which year. One school has a market share of about 70 per cent and the other schools all have market shares below 10 per cent. The HHI of Varberg is about 0.5, contingent on which year we consider. In the full sample, 9 per cent of the municipalities belong to the treatment group, which I explain in the next section. The average number of students and schools in a municipality is 1456 and 3.48 respectively.

Table 2: Municipality level descriptive statistics

All municipalities					
	HHI	Schools	Students	Treated	Monopoly
Mean	0.74	3.48	1455.72	0.09	0.48

Non-monopoly municipalities				
	HHI	School	Students	Treated
Mean	0.49	5.82	2420.53	0.17

Table 3 presents the descriptive statistics on the level of school. The average market share is 0.27 and schools face on average an HHI of 0.36. The average number of students is 402 and slightly above half of all schools are independent schools.

Table 3: School level descriptive statistics

All schools				
	Market share	HHI	Students	Private
Mean	0.27	0.36	401.87	0.52

4.2 Empirical framework

The potential endogeneity of competition makes estimation of the causal effect of competition on grade inflation problematic. A high grade level in the municipality makes it hard for an entrant to attract students and gain market share. If extensive grade inflation makes a market less attractive to entrants, grade inflation has a negative effect on competition. If there are simultaneous effects between competition and grade inflation, OLS estimation is not reliable for causal inference.

Another potential source of endogeneity is that there are unobserved variables determining both grade inflation and competition. An example of an unobserved variable is parental pressure on the teacher. While some parents pressure the teacher of their child to give higher grades, some do not. The degree of parental pressure is likely to also affect the degree of competition. If parents are dedicated, the market is probably less rigid and more attractive to entrants. If municipalities with dedicated parents attract more new schools, the dedication of parents triggers competition. Because the degree of parental pressure on the teacher is affects both grade inflation and competition, omitting parental pressure from the estimated equation makes the OLS estimator dependent on the effects of parental pressure on competition and grade inflation.

I cannot make causal inference if I do not solve the endogeneity problem of competition. The model I estimate is,

$$G_{ijt} = \alpha + \beta H_{jt} + \epsilon_{it} \tag{8}$$

G is the measured grade inflation in school i in municipality j at time t, α is a constant and H is the market concentration (HHI) in municipality j at time t. ϵ is a error term containing unobserved factors related to grade inflation in school i at time t. The parameter β is the effect of the HHI on grade inflation. If the HHI is related to variables in the error term, or grade inflation affects the HHI, the HHI is endogenous. As the HHI is endogenous, the estimate of β is biased.

To be able to make causal inference I use an Instrumental Variable (IV) approach. I use the shutdown of the schools belonging to the John Bauer Group (John Bauergymnasiet) in 2013 as an instrument for market concentration. Because of suspicious financial operations and governance in the John Bauer Group, the company went bankrupt and shut the schools down.[1] Because the shutdown was due to bankruptcy, I assume that the shutdown of these schools does not affect grade inflation, except through the impact on market concentration. The bankruptcy of a school ought not to affect grade inflation in other schools, except through the effect of the change in competition. If the shutdown's effect on grade inflation only runs through the change in competition, instrumenting the HHI with the John Bauer shutdown enables causal inference.

Not all John Bauer schools were actually shutdown in the bankruptcy. Some John Bauer schools were overtaken by other education providers and were therefore kept open. The municipalities that are regarded as treated by the shutdown are only the municipalities where the John Bauer school was shut down, yielding 17 treated municipalities. I refer to treated schools as schools that are located in these 17 municipalities. Hence, the treated schools are treated in the sense that their degree of competition is exogenously changed by the John Bauer bankruptcy.

To show that there was no particular selection among the John Bauer schools when some were overtaken, I compare the closed schools with the schools that were kept open in a regression analysis. Table 9 (see appendix) presents the results from the regression. Neither market share nor market concentration differs

[1] Swedish newspapers audited the financial operations and the shutdown. See for example Svenska Dagbladet Näringsliv (2013) and Dagens Nyheter Ekonomi (2013).

between the schools that were shut down and those kept open with statistical significance. Hence, it was other factors that decided which schools that were overtaken and kept open. The total number of students in the municipality (in thousands) is significant at the 10 per cent level but with a coefficient of insignificant magnitude.

Using the John Bauer shutdown as an instrument for competition, I follow the Two-Stage Least Squares (TSLS) procedure when estimating the effect of competition on grade inflation. The first stage measures the effect of the instrument on competition. The estimated effect of the instrument on competition is then used in the second stage to estimate the causal effect of competition on grade inflation. I hereby provide a detailed explanation of the procedure along with the assumptions needed.

The first stage measures the effect of the John Bauer shutdown on market concentration (the HHI). Because this is a difference-in-difference estimator (DID), it is important that the change in the HHI in the control group, i.e. untreated municipalities, is valid as counterfactual for the treated municipalities. Therefore, I must assume that if the John Bauer schools had not shut down, the HHI would change similarly in treated and untreated municipalities.

Figure 2: Common trend

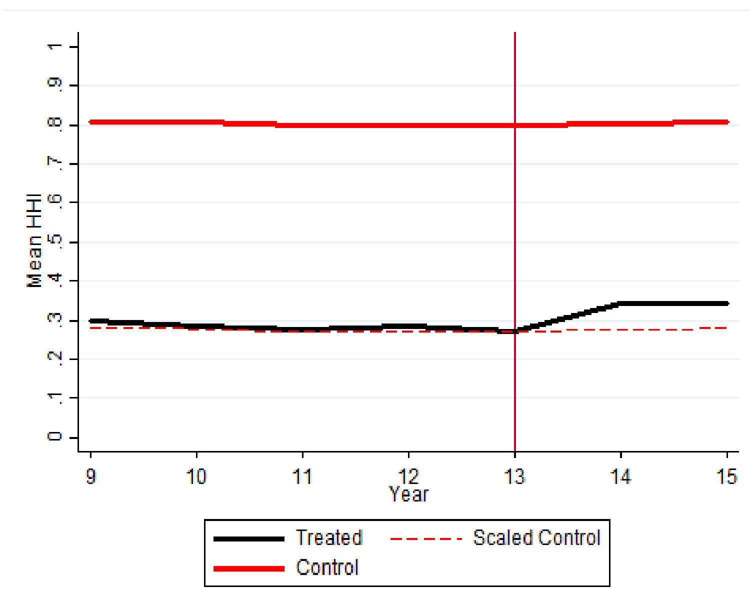


Figure 2 plots the mean HHI over time for the treated and untreated municipalities, respectively. For this analysis, I have collected extra data on enrollment from 2009-2012. If the trends are very similar before treatment, they would most likely continue to be in absence of treatment. The concentrations do trend marginally downwards in the pre-treatment period for both groups. However the trend is trivial. The important lesson is that the trends are almost identical prior to treatment. Therefore, I include a rescaled trend for the untreated municipalities to show how the trend of the treated municipalities would look in absence of treatment, i.e. the counterfactual. After year 2013, the treated municipalities experience a significant increase in concentration. The difference between the trend of the treated and the assumed counterfactual is interpreted as the causal effect of the John Bauer shutdown on competition in the treated municipalities.

Formally, regression analysis estimates the first stage. The estimated equation has the following expression,

$$H_{jt} = \alpha + \pi_{11} z_j + \pi_{12} T_t + \pi_{13} z_j * T_t + \epsilon_{it} \tag{9}$$

In this equation, market concentration H in municipality j in time t is the dependent variable. On the right hand side enter a constant α and a dummy for treatment group z, which is defined on the municipality level. T is a post time of treatment dummy and the last term is the interaction between z and T, which makes π_{13} the DID estimator. Because of conditioning on treatment group and time, the DID estimator is the effect of the John Bauer shutdown on the market concentration in the treated municipalities, i.e. an average treatment effect on the treated.

There are two techniques available for estimation of the effect of competition on grade inflation: TSLS and the Wald estimator. The Wald estimator is mathematically the same as the TSLS (Angrist and Pischke 2009). However, because it does not provide any standard errors, I also perform the formal IV estimation with TSLS. Both methods use the reduced form:

$$G_{ijt} = \alpha + \pi_{21} z_j + \pi_{22} T_t + \pi_{23} z_j * T_t + \epsilon_{it} \tag{10}$$

The parameter π_{23} in the reduced form estimates the effect of the John Bauer shutdown on grade inflation. This estimator is also a DID estimator. Because I assume that the effect of the shutdown on grade inflation only runs through the change in competition, combining the shutdown's effect on grade inflation with how much it changes competition gives the causal effect of competition on grade inflation. In calculation, the estimate from the reduced form is divided by the estimate from the first stage to create the Wald estimate:

$$\frac{\text{REDUCED FORM}}{\text{FIRST STAGE}} = \frac{\pi_{23}}{\pi_{13}} = \text{WALD} \tag{11}$$

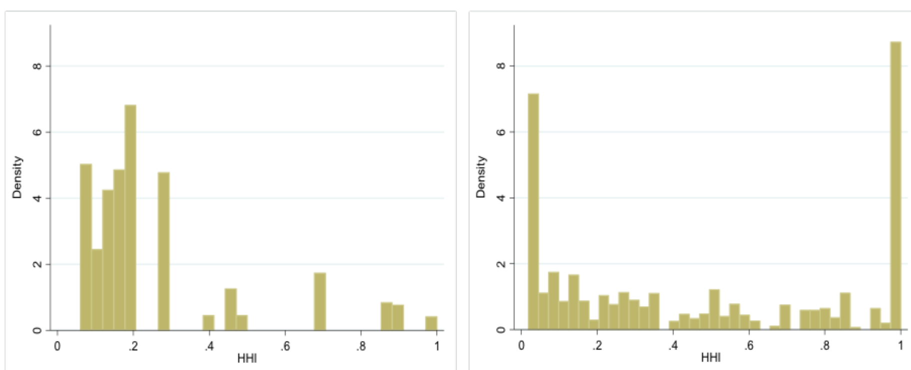
This estimate is the same as the IV estimate from TSLS, but it does not provide any standard errors. To pursue the formal TSLS procedure, I estimate the following equation,

$$G_{ijt} = \alpha + \pi_{31} z_j + \pi_{32} T_t + \beta H_{jt} + \epsilon_{it} \tag{12}$$

In this model, market concentration H is instrumented with $z_j * T_t$. Therefore, β is here the IV estimate of the effect of competition on grade inflation. The estimated coefficient should be the same as the Wald estimate.

The IV estimates the local average treatment effect (LATE). The TSLS estimation uses the variation in competition that the instrument causes. Because the effect from the instrument is exogenous, the IV estimates the causal effect of competition on grade inflation. However, schools do not perceive a shutdown of a school similarly. Some schools interpret it as an increased degree competition (compliers). However, because the HHI is an imperfect measure of competition, an increase does not necessarily mean that all schools competes fiercer. The IV estimates the average effect of competition on grade inflation for the schools that interprets the shutdown of a school as an increase in competition, i.e. the compliers. Therefore, it is called the local average treatment effect.

Whether competition affects grade inflation similarly in the untreated municipalities is a matter for discussion. Because this method estimates the LATE, I cannot make inference for untreated municipalities without assuming similarity. The untreated municipalities are on average more concentrated. However, because there are many monopolies driving up the average concentration, many untreated municipalities are much less concentrated than the average.



(a) Treated

(b) Untreated

Figure 3: Distribution of schools

Figure 3 enables comparison between the distributions of treated and untreated schools in year 2014. Recall that treated means affected by the John Bauer bankruptcy. Figure 3a shows that most of the treated schools are located in less concentrated municipalities. However, there are also some treated schools in more concentrated municipalities. Figure 3b shows that the untreated schools are more evenly distributed, but disregarding the many monopolies, the majority of the untreated schools are also located in less concentrated municipalities. Because the majority of both treated and untreated schools are located in less concentrated municipalities, they are similar in terms of exposure to competition. If they are also similar in other aspects, the estimated effect is likely to also apply for untreated schools.

For robust and reliable estimation, I want to control for as many other factors as possible. I use school level fixed effects to control for unobserved time invariant variables. As suggested, there might be unobserved factors that affects both competition and grade inflation. I use the example from before, parental pressure, to show what this means for the estimation.

$$G_{ijt} = \alpha + \pi_{31} z_j + \pi_{32} T_t + \beta H_{jt} + \delta PP_i + \epsilon_{it} \tag{13}$$

Suppose this is the true equation determining grade inflation. It is the same as Equation 12 but including parental pressure. Parental pressure is only dependent on which school considered, not time. Fixed effects on school level means that I recalculate each value as the deviation from the mean of the variable within each school:

$$\begin{aligned}(G_{ijt} - \overline{G}_{ij}) = (\alpha - \overline{\alpha}) + \pi_{41}(z_j - \overline{z}_j) + \pi_{42}(T_t - \overline{T}) + \beta_w(H_{jt} - \overline{H}_j) \\ + \delta'(PP_i - \overline{PP}_i) + (\epsilon_{it} - \overline{\epsilon}_i)\end{aligned} \tag{14}$$

Because a time invariant variable is does not change within the school, the difference between the variable value and its mean is always zero. Therefore, the time invariant variables are omitted from the equation:

$$\Delta G_{ijt} = \pi_{42} \Delta T_t + \beta_w \Delta H_{jt} + \Delta \epsilon_{it} \tag{15}$$

Within-estimation eliminates the omitted variable bias and controls for unobserved factors. Because using school fixed effects uses only variation within each school, it is called within-estimation. All variables that do not change over time are omitted and controlled for. If the variables that impact both competition and grade inflation, e.g. parental pressure, do not vary over time (and there are no other sources of endogeneity), the within-estimator has a causal interpretation. Therefore, using fixed effects supports the causal interpretation of the IV estimate.

The cost of fixed effects is reduced variation. The variation in competition is much less within schools than across. Therefore, within-estimation reduces statistical power. Using observations over many years increases variation and statistical power. However, it is harder to assume omitted variables do not vary when the time period is long. The choice of time length is therefore a trade off between variation and bias removal.

I allow for autocorrelation within schools. The unobserved variables that do change over time, i.e. not controlled for by fixed effects, are probably serially correlated. I assume that they are not correlated across schools, but allow for serial correlation within schools by using clustered standard errors on school level. This assumption is relaxed in the appendix, allowing for serial correlation also across schools in the same municipality.

4.3 Results

Table 4 presents results from the OLS estimation. While columns (1)-(3) present estimates without fixed effects, columns (4)-(6) present within-estimates. The dependent variable in columns (1) and (4) is the net share of students that receives a higher grade than test score (grade inflation). The columns (2) and (5), (3) and (6) have only the share of students that receive higher or lower, respectively. The decomposition of the measure of grade inflation enables me to investigate which of the deviations from the test score that drives the results.

If I omit a variable that affects both the HHI and grade inflation, the OLS estimate of the HHI is biased. The effect of the omitted variable on grade inflation the HHI determines in which way the OLS estimate is biased. In the example of parental pressure, I assume that parental pressure increases grade inflation but decreases the HHI (increases competition). Because the effects are of opposite directions, the OLS estimate is downward biased. That the within-estimate in column (4) is higher than the OLS estimate in column (1) suggests that the OLS is downward biased.

Except for omitted variable bias, there are probably simultaneous effects between grade inflation and competition. The simultaneity problem explains the positive coefficient of the HHI. The positive coefficient suggests that lower competition causes grade inflation to increase, which is not suggested by the literature or the model. Because fixed effects cannot account for simultaneous effects, the within-estimate is still unreliable and I need the IV estimate from TSLS.

Turning to the TSLS procedure, Table 5 presents the estimates of the first stage. Because the treatment group dummy does not vary over time, the coefficient is not identified. The time dummy (post treatment) is significant and positive. Hence, there is a general increase in concentration over time. However, the magnitude is very small. The interaction, i.e. the DID estimate, is also significant and positive. Therefore, the John Bauer shutdown caused market concentration to increase significantly more than it would in absence of the shutdown. The positive effect of the John Bauer shutdown on market concentration is expected (see Figure 2). The F-statistic of 16.95 excludes weak identification of the instrument. The TSLS estimation uses this variation, i.e. the variation in market concentration that the John Bauer shutdown causes, to measure the causal effect of competition on grade inflation.

Table 4: OLS estimates

VARIABLES	(1) Net Higher	(2) Higher	(3) Lower	(4) Net Higher	(5) Higher	(6) Lower
HHI	0.007	0.001	-0.006*	0.183***	0.215***	0.032
	(0.010)	(0.009)	(0.003)	(0.061)	(0.057)	(0.023)
Constant	0.183***	0.253***	0.070***	0.110***	0.165***	0.055***
	(0.005)	(0.005)	(0.002)	(0.026)	(0.024)	(0.010)
Observations	19,543	19,543	19,543	19,543	19,543	19,543
R-squared	0.000	0.000	0.001	0.000	0.001	0.000
Fixed Effects	NO	NO	NO	YES	YES	YES
Number of school				872	872	872

Note: Cluster standard errors (school level) in parentheses. The constant in within-estimation is the average fixed effect. P-values: *** = 1 %, ** = 5 %, * = 10 %.

Table 5: First stage

VARIABLES	(1) HHI
Treated	-
Post treatment	0.006***
	(0.002)
Treated*Post treatment	0.027***
	(0.007)
Constant	0.407***
	(0.002)
Observations	19,543
Number of school	872
R-squared	0.088
Fixed Effects	YES
F-stat	16.95

Note: Cluster standard errors (school level) in parentheses. The constant is the average fixed effect. P-values: *** = 1 %, ** = 5 %, * = 10 %.

Table 6 presents the estimates from the reduced form. I again provide the decomposed measures of grade inflation in columns (2)-(3) along with the standard measure, i.e. the net share that receives a higher grade than test score, in column (1). The treatment group dummy is again omitted due to within-estimation. We first consider column (1). The time dummy is significant and positive, suggesting a general increase in grade inflation over time. The coefficient of the instrument, i.e. the John Bauer shutdown, is significant and negative. Because the magnitude is much smaller than the general time trend, grade inflation still increases in the treated municipalities. However, the negative coefficient tells that the schools in the treated municipalities increase grade inflation significantly less than they would in absence of the shutdown. The negative coefficient means that the shutdown has a negative effect on grade inflation. Because the shutdown only affects grade inflation through the change in competition, this DID estimate is a causal effect of this specific decrease in competition on grade inflation.

Comparison between the columns shows which deviation from the test score that drives the effect. For the general time trend, coefficients are positive in both column (2) and column (3). The general time effect is positive in column (1) because the coefficient in column (2) is higher than in column (3). The dif-

Table 6: Reduced form

VARIABLES	(1) Net Higher	(2) Higher	(3) Lower
Treated	-	-	-
Post treatment	0.039***	0.047***	0.008***
	(0.005)	(0.004)	(0.002)
Treated*Post treatment	-0.019**	-0.012	0.007**
	(0.009)	(0.007)	(0.004)
Constant	0.163***	0.224***	0.061***
	(0.003)	(0.002)	(0.001)
Observations	19,543	19,543	19,543
R-squared	0.005	0.012	0.003
Number of school	872	872	872
Fixed Effects	YES	YES	YES

Note: Cluster standard errors (school level) in parentheses. The constant is the average fixed effect. P-values: *** = 1 %, ** = 5 %, * = 10 %.

ference in magnitude means that the share that receives higher grades increases more than the share that receives lower grades. Hence, the net share increases. The parameter of interest, i.e. the DID estimate, is only significant in columns (1) and (3). The positive coefficient in column (3) means that when the John Bauer schools shut down, the share that receives lower grades increases more than it would in absence of the shutdown. However, the share that receives higher grades is not affected by the shutdown. Therefore, the negative effect of the shutdown on grade inflation is driven by the increase in the share of students that receives lower grades than test scores.

Inflating a grade is easier if the student is lucky on the national test. Some students work very hard during the course, but scores low on the test. The teacher can inflate the grade by setting it above the score, claiming that the student performed much better in classroom assessments. It is true for some students, but roughly as many students should be lucky on the test, scoring above their true achievement level. If the teacher is honest, the students that score above the true level should be given lower grades, representing their true achievement level. However, because of the high test score, high grades for these students are easy to justify. Because it is the increase in the share that receives lower grades that drives the overall effect of competition on grade inflation, the grade inflation for the students scoring above their true achievement is most responsive to competition. Probably, the students that are lucky on the test have not put much effort during the course. Cutting class, omitting homework and signaling low interest is probably not appreciated by the teacher. Therefore, the teacher do not prioritize inflating these students' grades. However, the expected punishment cost is much lower if the test score is high. This is likely why the shares of lower grades are very low in general. If the teacher can easily state that a student performs well, she can easily inflate the grade. Therefore, the teacher can adjust the degree of grade inflation for the students who over performs on the national test without changing her risk of detection. When competition is suddenly lowered, e.g. due to the John Bauer shutdown, the teachers primarily reduce grade inflation for the students that over performs at the national test. This is exactly what we see in the estimate in column (3). The share of students that receives lower grades than test scores increases in response to the John Bauer shutdown.

When both of the first stage and reduced form are estimated, I calculate the Wald estimator using Equation 11. The Wald estimate is:

$$\frac{-0.0185}{0.0268} = -0.69$$

I here use the instrument's effect on grade inflation and competition, respectively, to estimate the causal effect of competition on grade inflation. As the coefficient is negative, market concentration decreases grade inflation. Because market concentration is inversely related to competition, the estimate shows that competition increases grade inflation. The magnitude of 0.69 is not trivial as it means that a decrease in market concentration by 0.1 in HHI corresponds to an increase of 0.069 in the net share of inflated grades. Because some test scores are inflated as well, the effect is probably underestimated. However, this estimator does not provide any standard errors to determine statistical significance. Therefore, I proceed the analysis with the TSLS results.

The IV estimates from TSLS are presented in Table 7. This table also includes two additional columns, (2)-(3), to give inference on which type of grade inflation that drives the results. The coefficient for the HHI in column (1) shows the estimated causal effect of market concentration on grade inflation. We already know the coefficient from the Wald estimate, but the TSLS estimation shows that it is statistically significant. From column (3) we observe again that the effect is driven by an increase in the share of students that receive lower grades than test scores.

I hereby provide an example to illustrate the coefficient's magnitude. Consider the comparison of two municipalities of roughly the same size, Falun (2992 students in 2015) and Östersund (2979 students in 2015). In Falun, there are 8 schools and the HHI is 0.19. Competition in Östersund is less fierce as there are 6 schools and the HHI is 0.3. Suppose a school suddenly shuts down in Falun, increasing market concentration by 0.11 to the same level as Östersund. All else constant, the causal effect from the decline in competition decreases the net share of students that receive higher grades than test score by 7.6 percentage

Table 7: TSLS estimates

VARIABLES	(1) Net Higher	(2) Higher	(3) Lower
HHI	-0.689*	-0.435	0.255*
	(0.389)	(0.298)	(0.146)
Treated	-	-	-
Post treatment	0.043***	0.049***	0.006**
	(0.006)	(0.005)	(0.003)
Constant	0.443***	0.401***	-0.042
	(0.158)	(0.121)	(0.059)
Observations	19,543	19,543	19,543
Number of school	872	872	872
Fixed Effects	YES	YES	YES

Note: Cluster standard errors (school level) in parentheses. The constant is the average fixed effect. P-values: *** = 1 %, ** = 5 %, * = 10 %.

points. Because the average net share of higher grades in Falun in 2015 is 0.26, this reduction in grade inflation is of 25.8 per cent.

The IV and OLS estimates are of different sign. The OLS estimate in Table 4 suggests a negative effect of competition on grade inflation. The IV estimate in Table 7 suggests a positive effect. The simultaneous effects between competition and grade inflation are probably the reason. If the negative effect of grade inflation on competition dominates the effect of competition on grade inflation, the simple OLS estimate provides a positive coefficient. Because the TSLS exploits exogenous variation in competition, the IV estimate reveals the causal effect from competition on grade inflation. Incorrectly assuming exogeneity can lead the researcher to the wrong conclusion. Therefore, it is important to treat potential endogeneity properly. Results from simple estimations that do not address endogeneity must be interpreted very carefully.

5 Characteristics of severe grade inflators

In section 4, I show the causal effect of competition on grade inflation. However, the literature suggest that grade inflation decreases student effort and results. In this section, I exploit the suggested effect of grade inflation on learning outcomes to identify severe grade inflators. I use the identification to evaluate what characteristics correspond to grade inflation.

I use the grade inflation on the first educational level (of upper secondary school) and the educational development over time to identify severe grade inflation. To measure educational development over time, I pool observations within subjects and schools. Therefore, each school is divided into three observations, one for English, one for Swedish and one for Mathematics. I define the average grade inflation on the first level for each department (subject) as the net share of students that receive higher grades than test scores on the first level. I define educational development by calculating the change in average test score between the first and second level. If the difference is negative, educational achievements decline in the department. I classify severe grade inflation if the department conducts grade inflation on the first level and educational achievements decline.

Figure 4: Severe grade inflation

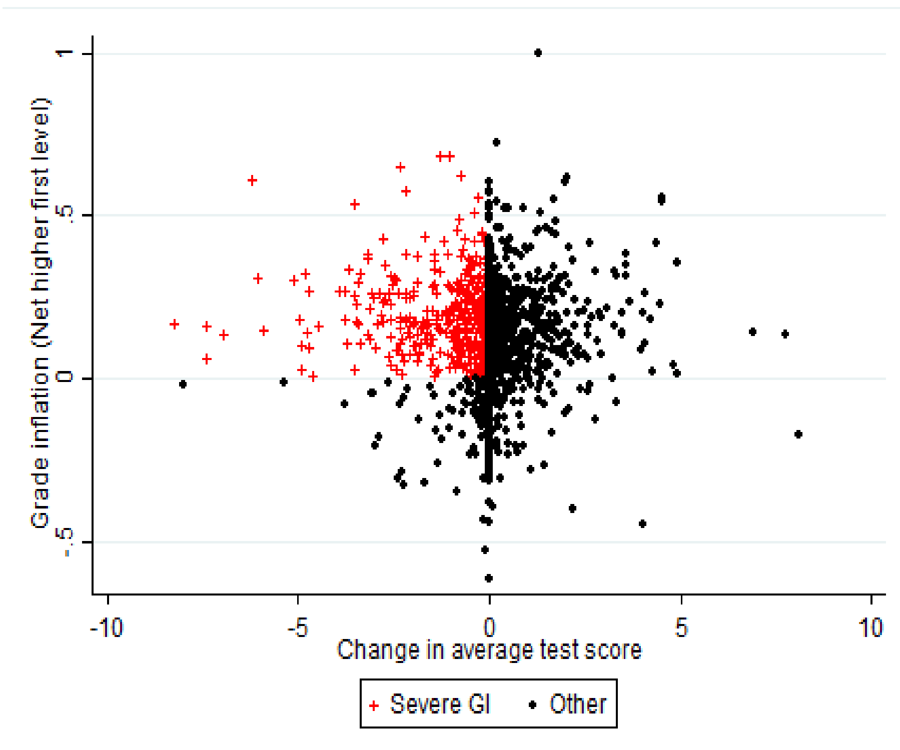


Figure 4 plots grade inflation on the first level on the change in average test score. The departments classified severe grade inflators are plotted with red plus signs (upper left quadrant). The other departments are plotted as black dots. Because majority of departments do inflate grades on the first level, most of the observations are above 0 on the y-axis. However, as students respond differently to grade inflation, not all results decline. Some departments manage to improve educational achievements despite the grade inflation they conduct. The departments that inflate grades and where results also decline, i.e. the severe grade inflators, inflate grades in a manner that induces students to learn less. It is therefore not said that they inflate grades the most, just that the consequences (for the students) are more severe in these departments.

I use regression analysis to investigate which characteristics do correspond to severe grade inflation. With probit and OLS regressions, I estimate conditional correlations between severe grade inflation and different characteristics that possibly relates to grade inflation.

Table 8 presents the results from the regression analysis. I perform two estimations. Column (1) provides the estimates from a non-linear estimation (probit) and column (2) from OLS. Because the dependent variable is a dummy and coefficients are interpreted in terms of probabilities, linear estimation is not really intuitive. It is not intuitive because marginal effects are probably not constant. The probit estimation allows for non-linear marginal effects, but requires distributional assumptions. However, exact estimation of marginal effects is not the purpose of this analysis. Because I acknowledge that the measure of severe grade inflation is somewhat coarse, the purpose of this analysis is merely statistical significance and approximate magnitude. Therefore, I use both estimations to draw conclusions. The reported coefficients from the probit estimation in column (1) are marginal effects, evaluated at the mean of the respective covariate.

Market concentration corresponds to severe grade inflation. As the results of the empirical analysis in the previous section shows, market concentration has a negative relationship with severe grade inflation. The variable 100 students per school is an alternative measure of market concentration and its coefficient is positive. However, it is of negligible magnitude.

School size is positively related to grade inflation. Wikström and Wikström (2005) suggests that teachers in large schools have less incentives to inflate grades. However, these results tell a different story. Both market share and school size (100 students in school) are positively related to severe grade inflation in this analysis. In Section 3, I suggest that the risk of detection decreases when the teacher can "hide in the crowd". This effect makes school size and market share positively related to severe grade inflation. Teachers on large schools still have the incentive to inflate grades, but the expected cost of punishment is smaller.

The level of education among the teachers on the department does not correspond to severe grade inflation. In Section 3, I suggest that the morale of teachers should have a negative impact on grade inflation. However, I also claim that the effect from moral cost of grade inflation is only a fraction of the effect from expected punishment and unemployment risk. It is reasonable to assume that a more educated teacher has higher professional conduct standards, i.e. morale. The coefficient for teacher education in this analysis is not statistically significant. Therefore, if morale strongly correlates to the level of education, morale has no effect on grade inflation.

Table 8: Determinants of severe grade inflation

VARIABLES	(1) Severe GI	(2) Severe GI
HHI	-0.116*	-0.010*
	(0.066)	(0.054)
Market share	0.133**	0.127**
	(0.059)	(0.050)
Teachers' education	0.070	0.051
	(0.047)	(0.041)
Students per teacher	-0.000	-0.000***
	(0.002)	(0.000)
Mathematics	0.146***	0.135***
	(0.021)	(0.018)
Swedish	0.080***	0.072***
	(0.019)	(0.015)
100 students in school	0.008***	0.010***
	(0.003)	(0.003)
Schools in municipality	-0.002	-0.001
	(0.003)	(0.003)
Average test score	0.008*	0.008**
	(0.004)	(0.004)
100 students per school	0.000*	0.000**
	(0.000)	(0.000)
Students in municipality	0.000	0.000
	(0.000)	(0.000)
Constant		-0.112**
		(0.055)
Observations	2,471	2,471
R-squared		0.065

Note: Robust standard errors in parenthesis. P-values: *** = 1 %, ** = 5 %, * = 10 %.

Grade inflation depends on the subject. These results suggest that severe grade inflation is most common in Mathematics, secondly in Swedish and the least in English. A difference is possible, but a more likely interpretation is that it is actually the test scores that are inflated differently depending on the subject. Because national tests in Swedish and English include writing and oral performance, the teacher must interpret vague assessment directives much more than the teacher in Mathematics. Because the determination of test score in Mathematics is less diffuse, the teacher in Mathematics is given less room to inflate the score. Hence, a larger fraction of the grade inflation is visible when comparing the grade with the score.

Higher performing schools inflate grades more. The average score has a positive relationship with severe grade inflation. Schools with high performing students probably address the students that value high grades. If high grades are the school's primary means of competition, the incentive for inflating the grades is of course higher. This corresponds to the parameter θ in the theoretical model. If the parameter θ, i.e. importance of grades, increases, so do the incentives for grade inflation.

I am not able to show an effect of market size using the measure of number of students in the municipality.

The conclusions from this section relate mostly to the theoretical model I present in Section 3. I explore the mechanism suggested from and used in previous literature. The adverse effect of grade inflation on student effort allows me to construct a classification of severe grade inflation. I use the classification to evaluate what school and departmental characteristics that correspond to grade inflation, and I find support for the theoretical model.

The importance of grades relate positively to grade inflation, as the schools with high grades inflate grades more. The education level of teachers does not relate significantly to grade inflation. Because I assume that the education level relates to professional conduct, I interpret the absence of a relationship between the teacher's education and grade inflation as an indication that the cost of morale and bad conscience does not significantly determine grade inflation. Hence, the ϕ parameter is appropriately assumed to be very small and could be omitted.

The expected punishment cost decreases with competition. That competition relates positively to grade inflation does not alone prove that there is the hide-in-the-crowd effect that spurs grade inflation. However, the positive relationships between grade inflation and both school size and market share provide additional proposal for such effect. All together it suggests that the hide-in-the-crowd effect increases grade inflation when competition increases.

I also find significant differences between the subjects. Because there are no suggestions in the literature that these subjects ought to differ in grade inflation, I interpret these differences as differences in how much the national test scores are inflated. This supports the idea that the relationship between grade and test score probably underestimates the true grade inflation.

6 Conclusion

I show that competition has a positive effect on grade inflation. Because I use instrumental variable estimation and fixed effects, I solve the endogeneity problem of competition. Because of IV and fixed effects, the estimated effect has a causal interpretation.

The literature suggests this relationship but previous empirical results are vague. Most of the previous research on competition and grade inflation considers the market for elementary schooling. Because of differences in admission and school choice, the market for elementary schooling is likely more rigid than the market for upper secondary schooling. Therefore, the effect of competition on grade inflation is probably not as distinctive on the market for elementary schooling as it is on the market for upper secondary schooling.

The measure of grade inflation affects the results. I exploit the new grading system and the most appropriate measure of grade inflation to get highest precision possible. The main problem for this measure is that it underestimates the true grade inflation. This is probably also why Vlachos (2010) only finds a very small effect from competition on grade inflation. This problem causes my results to understate the true effect of competition on grade inflation.

How to measure competition is a difficult but important choice. Previous studies use more simple measures of competition. I use a more detailed and informative measure, suggested by the literature. This measure is the Herfindahl-Hirschmann Index, and it allows me to accurately quantify the degree of competition. With an accurate measure of competition, I get high precision in my estimation of the effect on grade inflation.

The endogeneity of competition must be addressed for reliable estimation. Comparison between the IV and the OLS estimates teaches an important lesson. If the problem of simultaneous effects is not addressed properly, the researcher can draw the wrong conclusion. Because the OLS only estimates conditional correlations, simultaneous effects make causal inference inappropriate. Within-estimation does not alone solve this problem. If I do not use the IV approach I mistakenly, with this sample, believe that competition decreases grade inflation.

The exploration of a well-established mechanism supports the theoretical model I present. Theory and previous findings suggest that grade inflation leads to declining learning outcomes. I exploit this mechanism and estimate what school and municipal characteristics that relates to severe grade inflation. My findings support the theoretical model I provide. It also provides further assurance that the estimated effect of competition on grade inflation is underestimated.

I am able to draw some conclusions about how competition causes grade inflation in more detail. The findings show that grade inflation for students that score high on the test but do not achieve much during the course responds to competition. The grade inflation for students that score low on the test but put much effort during the course does not respond to competition. That grade inflation for over performers on the national test responds more to competition indicates that the expected punishment cost is important. This further investigation of exactly how grade inflation is conducted is an appropriate question for further research. Except for competition, what makes schools and teachers more likely to inflate grades is for future research to explore.

In my theoretical model, I also suggest that competition decreases true educational quality. Because of grade inflation, using grades as a measure of educational quality is truly misleading. Previous studies of the effect of competition on educational quality (using grades as measure of educational quality) that does not control for inflated grades cannot be relied upon. Now that the lesson of competition and grade inflation is learned, it is possible to proper investigate the effect of competition on educational quality.

References

Ahlin, Å. (2003). Does school competition matter? Effects of a large scale school choice reform on student performance. *Working Paper*, 2003:2, Department of Economics, Uppsala University.

Ahlin, Å. & Johansson, E. (2000). Demand for local public schooling: Another brick in the wall. *Working paper*, 2000:12, Department of Economics, Uppsala University.

Akerlof, G. A. (1979). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500.

Angrist, J. D. & Pischke, J. (2009). *Mostly harmless econometrics: an empiricist's companion.* Princeton: Princeton University Press

Betts, J. R. & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22, 343- 352.

Bonesrönning, H. (2004). Can effective teacher behaviour be identified? *Economics of Education Review*, 23, 237-247.

Böhlmark, A. & Lindahl, M. (2007). The impact of school choice on pupil achievement, segregation, and costs: Swedish evidence. *IZA DP* 2786.

Costley, C. K. (2014). *Stop giving in to higher grades: Ten suggestions on how to fight grade inflation.* Arkansas Tech University: Russelville.

Dagens Nyheter Ekonomi. (2013). John Bauer-elever snuvades på 200 miljoner. *Dagens Nyheter*, 24 October [Online] Available at: http://www.dn.se/ekonomi/john-bauer-eleversnuvades-pa-200-miljoner/ [Accessed 10 May 2016].

Figlio, D. & Loeb, S. (2010). School accountability. In *Handbook of the economics of education*, 3. Elsevier B.V.

Hoxby, C. (1994). Do private schools provide competition for public schools? *NBER WP* 4978, Cambridge.

Hsieh, C. & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90, 1477-1503.

Jacob, B. & Levitt, S. (2003). Rotten apples: an investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3), 843-877.

Sandström, F. M. & Bergström, F. (2005). School vouchers in practice: competition will not hurt you. *Journal of Public Economics*, 89, 351-380.

Svenska Dagbladet Näringsliv. (2013). En miljard fattas i John Bauer-konkursen. *Svenska Dagbladet*, 5 August [Online] Available at: http://www.svd.se/en-miljard-fattas-i-john-bauerkonkursen [Accessed 10 May 2016].

Walsh, P. (2010). Does competition in schools increase grade inflation? *Education Working Paper Archive.*

Wikström, C. & Wikström, M. (2005). Grade inflation and school competition: an empirical analysis based on Swedish upper secondary schools. *Economics of Education Review*, 24, 309-322.

Vlachos, J. (2010). *Betygets värde: en analys av hur konkurrens påverkar* *betygssättningen vid svenska skolor*. Uppdragsforskningsrapport: 2010:6. Konkurrensverket: Stockholm.

Vlachos, J. (2016). Hur ser betygsinflation ut? *Economistas*, 27 April [Online] Available at: https://ekonomistas.se/2016/04/27/hur-ser-betygsinflation-ut/ [Accessed 11 May 2016].

Appendix

Table 9: John Bauer comparison

VARIABLES	Shut
Students in school	0.001
	(0.001)
Market share	-3.992
	(3.789)
HHI	0.440
	(0.850)
100 students in municipality	-0.004*
	(0.002)
Constant	0.780***
	(0.271)
Observations	30
R-squared	0.096
F-stat	4.185

Note: Robust standard errors in parenthesis.
P-values: *** = 1 %, ** = 5 %, * = 10 %.

...

For robustness check, I perform the TSLS procedure but with errors clustered on municipality level instead. Clustering on municipalities allows for serially correlated errors across schools within the same municipality. An obvious problem is that there are only 229 clusters, which is much less than if I cluster on schools where I have 872 clusters. The decrease in clusters reduces statistical power and yields less convincing results. In Table 10, I present results from the four regressions: (1) − OLS, (2) − First stage, (3) − Reduced form, (4) − TSLS. Fixed effects on school level are used in each estimate.

Statistical significance maintains except in the final IV estimation. Because the F-statistic in the first stage is now only 7.815, the instrument suffers from weak identification. However, the reduced form is still significant. Therefore, the DID interpretation of the effect of the John Bauer shutdown on grade inflation is still valid, indicating that even when relaxing the assumption of errors being uncorrelated across schools, a decrease in competition causes grade inflation to decrease. The small number of clusters and the weak identification of the instrument reduce the statistical power enough for the IV estimate to lose its significance.

I assume errors are serially correlated within but not across schools in the paper because there is no strong argument against. It is very likely that grade

Table 10: Clustering on municipality level

VARIABLES	(1) Net Higher	(2) HHI	(3) Net Higher	(4) Net Higher
Treated		-	-	-
Post treatment		0.006*	0.039***	0.043***
		(0.003)	(0.005)	(0.007)
Treated*Post treatment		0.027***	-0.019*	
		(0.010)	(0.011)	
HHI	0.183***			-0.689
	(0.061)			(0.535)
Constant	0.110***	0.407***	0.163***	0.443**
	(0.025)	(0.002)	(0.003)	(0.218)
Observations	19,543	19,543	19,543	19,543
R-squared	0.000	0.088	0.005	
Number of school	872	872	872	872
F-stat		7.815		

Note: Robust standard errors in parenthesis. P-values: *** = 1 %, ** = 5 %, * = 10 %.

inflation depends on other factors that are the same within a specific school, where teachers interact with each other every day. However, schools are very different even if they are placed in the same municipality. Therefore, the step from clustering on schools to clustering on municipalities is not small and should only be taken for good reasons. Because I do not find good reasons, I cluster on schools.

As a robustness check, it does not invalidate the conclusion that competition causes grade inflation. Because the reduced form is still significant and the effect is probably underestimated, the IV estimate could survive the clustering on municipalities if there were more municipalities or time periods.