IFN Working Paper No. 1217, 2018

Trust-Based Evaluation in a Market-Oriented School System

Jonas Vlachos

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Trust-Based Evaluation in a Market-Oriented School System*

Jonas Vlachos♠

2018-05-22

Abstract

In Sweden, a trust-based system of school performance evaluation meets a market-oriented school system with liberal entry conditions for voucher-funded private providers. National standardized tests are graded at the local school and what ultimately matters to students are teacher-set grades. This paper finds that privately run free schools systematically set higher grades than public schools when controlling for their achievement on national tests. The differences between municipal and free schools are larger when more reliable tests are used to account for achievement. Differences in grading standards between providers are substantial and most of the performance advantage in teacher-set grades that free schools enjoy can be attributed to more generous grading. The results also indicate that different private providers do not necessarily respond symmetrically when faced with similar market conditions and act under the same regulatory regime.

Keywords: Private provision of public services, School performance evaluation, School vouchers

JEL codes: H44, I28, L51

* This paper is prepared to appear as a chapter in Dahlstedt, M. & Fejes, A (eds) *Neoliberalism and market forces in education: Lessons from Sweden*. London: Routledge. I am grateful for helpful comments by Anders Auer, Andreas Fejes, Kristian Koerselman, Christian Lundahl, Per Molander, Jonas Sandqvist, and Anna Sjögren. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

♠ Department of Economics, Stockholm University and IFN. E-mail: jonas.vlachos@ne.su.se.

1. Introduction

In Sweden, a trust-based evaluation system meets a market-oriented school system with liberal entry conditions to voucher funded private providers, several of which are for-profit corporations. The purpose of this chapter is to analyze how different providers respond to a system where teachers and schools are given a high degree of discretion with respect to school performance evaluation. While all providers are likely to somehow respond to the incentives provided by the system, the objectives differ across providers and their responses may therefore also differ. After all, one of the motivations for allowing private providers is precisely that they are assumed to be more sensitive to demand pressures. Such greater responsiveness can be beneficial, but it can also be difficult to contain when private and public objectives differ. Since the integrity of the testing and grading systems is a quintessential public good, the potential for a conflict of interest is obvious.

The analysis of the private provision of publicly funded education has a long tradition, not the least in economics. Perhaps most famously, Friedman (1962) argued in favor of a voucher system fully open to private providers. Subsequently, Shleifer (1998) and Hoxby (2003) have forcefully made the case for allowing private provision of elementary and secondary education. These conclusions are based on the notion that private providers are more sensitive to the demands of families and that the competition that these providers induce will improve the quality also among public providers. There are therefore two distinct parts of the argument; one based on choice and competition between schools, and one based on private providers being more responsive to the preferences of families.

If families are well informed and there are no discrepancies between the private and public perceptions of school quality, the case for market forces such as competition and private provision is strong. However, the assumption that families are well informed about school quality is restrictive and unlikely to hold in practice. After all, disentangling the impact of the school from that of its student composition is a challenge also for trained researchers. Perhaps even more questionable is the assumption that private and public notions of school quality are well aligned. Several chapters in this book highlight how school choice can result in segregation along various dimensions. While families may regard this as individually desirable, such segregation may not be in the general public interest. Schools may also compete by catering to students' preferences for various consumption amenities rather than by offering high-quality education.[1] Again, this can be unproblematic from an individual's viewpoint but may not be in the public interest.

An area in which both information problems and a conflict between private and public interests can arise is in school performance evaluation. Various metrics such as test scores and grade point averages (GPA) are used for performance evaluation, both formally by the authorities and informally by families. Test scores and GPA:s are often consequential for the students and both families and schools have incentives to game the evaluation system. How strong these incentives are depends on how the performance measures are used and the ease by which the system can be gamed.[2] In Sweden, the evaluation system is based on subject tests and national standardized tests that are locally graded according to centrally set criteria. What ultimately matters for the students are teacher-set grades.

The main purpose of the standardized tests is to help teachers conform to nationally uniform grading standards, but there is no formal connection between test results and grades. Teacher-set subject and

[1] A recent study of US college choices reveals that while most students value consumption amenities, only high-achieving students have a taste for academic quality. Colleges also respond to these demand pressures (Jacob et al., 2018).

[2] There is a large literature on how accountability systems are gamed. See Koretz (2017) for a recent contribution.

course grades are combined into a GPA which is the main selection mechanism when students move from compulsory to upper-secondary school, and from upper-secondary school to tertiary education. Apart from being highly consequential to the students, the GPA is the main performance metric used in various rankings. Hence, the Swedish system is highly trust-based, easy to game, and the incentives to game it are strong. The deeper question addressed therefore concerns the compatibility between trust and market oriented systems.

The focus of this chapter is on differences between providers and not on the broader impact of competition. The reason is partly methodological but mainly that such differences between providers are important in their own right. In what follows, I will present the system of standardized testing and grading in Sweden. This will provide the basis for the empirical methodology applied when analyzing differences in grading standards between providers. While the main analysis presented here is limited to the compulsory level, some previous findings that relate to the upper-secondary level will be discussed towards the end of the chapter.

2. The voucher system and the main corporate providers

The Swedish education system is comprised of nine years of compulsory schooling, followed by three years of upper-secondary education. While voluntary, 99 percent of the Swedish youth enroll in some type of upper-secondary program. Approximately 15 percent of all students enter various preparatory programs aiming at preparing non-eligible students for the 18 different three-year vocational or academic programs which the remaining 85 percent enroll in.

Subject to national regulation, Swedish municipalities are in charge of providing and maintaining the quality of both compulsory and upper-secondary education. Schools are financed by the municipalities and could either be run by the municipality itself or by an independent private provider. Both municipal and free schools are funded in relation to its number of students. By law, municipalities are required to compensate free schools in the same way as they fund municipal schools. In several municipalities the funding available to each school is adjusted for the socio-economic background of the school's students, but such systems vary substantially between municipalities.

The voucher system was introduced in Sweden in 1992 and coverage is universal at both the compulsory and the upper-secondary level. The national school inspectorate review free school applications based on the general educational plan, the premises to be used, the financial strength of the provider, and forecasts concerning the student population. There are few restrictions on who is allowed to own and manage a school. For example, no prior experience in education is required and for-profit, incorporated, schools are fully accepted. Prior to approval, municipalities are allowed to voice concerns regarding the impact of additional school entry. In practice, the concerns raised by municipalities appear to have had a limited impact on the approval rate of new voucher schools.[3]

When the voucher reform was initiated, only around one percent of students attended free schools. As of 2017, this share has increased to around 15 percent at the compulsory and 26 percent at the upper-secondary level. Out of the compulsory school students at free schools, 68 percent attend schools that are for-profit (limited liability corporations) and the corresponding fraction at the upper-secondary level is 86 percent.[4]

[3] Skolinspektionen (2013a) report that 15 percent of the rejections at the upper-secondary level and only a hand full of applications at the compulsory level were turned down due to municipal concerns.
[4] These number refer to the 2015/16 school year (Friskolornas Riksförbund, 2017).

Several larger multi-school corporations are active in the free school sector and much of the expansion has occurred within such corporations. At the compulsory level, the three largest actors are Academedia, Internationella Engelska Skolan (IES), and Kunskapsskolan.[5] Both Academedia and the IES are listed corporations with the shares publicly traded, while Kunskapsskolan is owned by the founder's holding company. In the data used in the main analysis, 55 schools belong to Academedia, 23 to the IES, and 28 to Kunskapsskolan.

The three groups have quite different profiles. Academedia organizes its compulsory schools under the *Pysslingen* and *Vittra* labels, and these schools differ both in terms of size and pedagogy. Academedia assists schools by providing quality management and according to the webpages the educational philosophy is based on the ideas behind Montessori and Reggio Emilia.[6] Academedia was founded in 1996 and has grown by partly by acquiring existing schools. In 2008, Academedia acquired Vittra which was established in 1992. Pysslingen started to run pre-schools already in the 1980's and it was acquired in 2011.

The IES was established in 1994 and it is currently the fastest growing corporation at the compulsory level. The IES is based on the idea of bilingual education and a substantive part of the teaching is in English. Apart from it bilingual profile, the IES has similarities to the US no-excuse charter schools with strict and consistent disciplinary standards, the concept of 'tough love', high academic expectations, and traditional classroom teaching as hallmarks. Kunskapsskolan was founded in 1999 and has a distinct profile and can be seen as the polar opposite of the IES. Its concept is based on personalized education catering to the individual goals and needs of the students. Traditional classroom teaching is downplayed and the teacher's role and a guide and mentor is stressed.

The rest of the free school sector at the compulsory level is varied. There are several other smaller multi-school corporations but also many stand-alone schools that are either incorporated or run as non-profit entities. In total, approximately 17 percent of students exit compulsory education from a free school. Eight percent of these free school students attend confessional schools, four percent Waldorf schools, and the rest attend schools with a general profile. Some of the general profile schools specialize in music, sports, or the arts, but no aggregate statistics on such specializations are available. In the analysis that follows, I will put all these (286) schools into one category. Needless to say, this aggregation hides a substantial amount of heterogeneity within the group.

3. Grading and testing

In the Swedish compulsory school, 17 subjects are taught and graded.[7] The top grade (A) awards the student 20 credits and the lowest pass grade (E) gives 10 credits. Grades between A and E are given in 2.5 credit intervals and a fail grade (F) is given zero credits. All subject grades count towards the final merit value which I here call the GPA. Grades are awarded by the teacher and should reflect the student's performance in relation to nationally set criteria. The GPA is the main selection mechanism when students apply to upper-secondary school and the grades are thus high-stakes for the students. To the schools, the school level GPA is one of the most important performance metrics.

To aid teachers in their grade setting, standardized test that are locally graded based on central grading directives are administered to all students in mathematics, Swedish, and English towards the

[5] The major commercial actors and their histories are briefly presented in Abrams (2016, chapter 11).
[6] See https://vittra.se/om/ and http://www.pysslingen.se/skolor/om-oss/academedia-modellen/.
[7] In the analysis, I will not make use of modern languages. The reason is that student have a choice of languages, meaning that the coverage of grading data is far from complete. This is because a minimum of ten students is required for the authorities to publish school level data.

end of the final year of compulsory school (grade 9). Each year, schools also take national tests in one of the sciences (biology, chemistry, or physics), and one of the social sciences (civics, geography, history, or religion). In the sciences and social sciences, the Swedish National Agency for Education decides which subject that is administered at each school with the aim of guaranteeing a representative sample. To the schools, the tests are formally low-stakes but in practice, but in particular the tests in the core subjects are regarded as important. A majority of students receive the same final grade as the grade on the test but the relation between test grades and final grades differs substantially between schools. The school-level results of these tests are available on government webpages and thus accessible for those interested. Although there is no formal alignment between the subject grades set by teachers and the test results, the National School Board reports deviations between subject grades and the test grades.[8]

The analysis that follows builds on the idea that deviations between grades and test results can be used to compare grading standards between schools. This approach is based on the premise that there is more teacher discretion in grade setting than in the grading of national tests. While this premise is reasonable, it is important to note the implications of tests being locally graded. The Swedish Schools Inspectorate has re-graded a large number of national tests using external examiners. The analysis of this effort reveals substantial differences in test grading standards between schools. Across subjects, external examiners tend to have stricter test grading standards than the original teacher but the discrepancy between external and internal examiners varies between subjects and test type (Skolinspektionen, 2013b). Unsurprisingly, there is a weaker alignment between external and internal assessments on tests with a high degree of teacher discretion. In mathematics, the discrepancy between external and internal assessment is smaller but not trivial.

An implication is that there is some degree of teacher discretion in the grading of all tests and test reliability is a matter of degree. Despite the caveat mentioned above, the mathematics test can be labelled as relatively reliable. In English, three different tests are administered and the correspondence between examiners is strong for the test in reading and listening comprehension (English B) but substantially weaker when assessing writing abilities (English C). No comparison has been done for the test of verbal ability (English A), but it is reasonable to assume a high degree of discretion for this test. In Swedish, the correspondence is again relatively strong for reading comprehension (Swedish B), albeit not as strong as for English B. The correspondence is weaker for the verbal test (Swedish A), and weakest for essay writing (Swedish C). In the sciences, the differences between external and internal evaluations are large. There has been no external re-grading of the tests in the social sciences but since these tests highly rely on open-ended questions, it is fair to characterize them as less reliable. It is further worth pointing out that free schools are on average more generous in their internal test grading, although the differences from public schools are not always statistically significant.[9]

The discussion of test reliability is summarized in Table 1. The most reliable tests are the English B test and the mathematics test, while Swedish B is most reliable among the three tests in Swedish. The other tests are all less reliable and for the aesthetic and practical subjects (arts, home economics, technology, music, crafts, and sports) there are no tests administered.

[8] In late 2017, the Swedish parliament approved of a bill stating that the national tests should be given a larger weight in the grade setting. If this has an impact on grade setting remains to be seen.

[9] Compared to public schools, free schools are also relatively generous in their test grading (see Hinnerich and Vlachos, 2013 and 2017).

Table 1. Comparison of test reliability

	More /Less reliable	Test grades (mean)	Test grades (sd)	Obs
Mathematics	More	11.58	2.06	5648
English A	Less	14.79	1.63	5648
English B	More	14.97	1.82	5648
English C	Less	14.17	1.74	5648
Swedish A	Less	14.71	1.47	5472
Swedish B	(More)	13.41	1.69	5472
Swedish C	Less	12.37	1.84	5472
Sciences	Less	12.15	2.37	5642[a)]
Social sciences	Less	12.81	2.19	5647[b)]
Aesthetic and practical	Only grades	14.26[c)]	1.62[c)]	33876[c)]

Note: The sciences are biology, chemistry, and physics. Social sciences are civics, geography, history, and religion. Aesthetics and practical are arts, home economics, technology, music, and craft. [a)] The number of subject grades in the sciences is 16944. [b)] The number of subject grades in the social sciences is 22592. [c)] Numbers refer to subject grades rather than test grades.

4. From test results to grading standards

At the school level, the test results should broadly reflect the level of student knowledge. Hence it is natural to consider a mapping from test results to the subject grades. A simple mapping would be to just compare averages, but this assumes that the relation is similar throughout the distribution of test results. Such an assumption is unlikely to hold in practice. First of all, the grading scale is not symmetric and more grade points are awarded for a move from fail (F) to pass (E) than for a move from B to A. Second, teachers are particularly prone to give a passing grade to students who have failed the test which is unsurprising given the grave consequences of a failing grade. Third, students who score an F on the test cannot be graded down, while students who score an A cannot be graded up. For all these reasons, there will be a tendency for schools with many low-achieving students to appear relatively lenient in their grading compared to schools with many high-achieving students when comparing averages. An additional complication is that test absenteeism is unlikely to be random.

When comparing grading standards across providers, I therefore run school level regressions of the below type using ordinary least squares. Subscript *i* stands for school, *s* for subject, and *t* for time period. Since tests are not given in all subjects, subject grades are sometimes mapped to a test in a different subject. Test subjects are therefore called $\hat{s}$. The fixed effects $\mu_{s,t}$ allow the intercepts to vary between subjects and years. The term $f_{s,t}\left(Test\ grade_{i,\hat{s},t}\right)$ is a flexible function of test grades and the relation between these and subject grades is allowed to vary by subject and year.[10] I account for test absenteeism by controlling for the share of test takers and again this relation is flexible across subjects and years. Finally, $\varepsilon_{i,s,t}$, is an error term.

$$Grade_{i,s,t} = \beta_j Provider_i^j + f_{s,t}\left(Test\ grade_{i,\hat{s},t}\right) + \gamma_{s,t} Share\ test\ takers_{i,\hat{s},t} + \mu_{s,t} + \varepsilon_{i,s,t}.$$

Schools are grouped depending on which provider group *j* they belong to. The null-hypotheses that are being tested are that the relations between test grades and subject grades are the same across providers, i.e. that the β_j:s are equal to zero (0). When testing these hypotheses, the standard errors are clustered at the school level to account for intra-school dependence across subjects and between

[10] I use a fourth degree polynomial of test grades but the results are not sensitive when even more flexible functional forms are used. See table A9 in the appendix.

time periods. Schools are grouped into municipal schools (reference category), Academedia, IES, Kunskapsskolan, and Other free school providers.

Based on the discussion of test reliability, I use the tests grades for mathematics, English B, and Swedish B when mapping the grades in each respective core subject. For the sciences and social sciences, I use either the own subject test or the test in mathematics. The mathematics test is used since the test is reliable and since performance in mathematics is strongly correlated to overall educational performance across subjects. Further, it is the advantage of being available for all schools and years, while the tests in the natural and social sciences are rotating and only administered every third or fourth year on average. The mathematics test is also used for aesthetic and practical subjects.

This brings me to the major caveat of the approach. A school in which the students for some reason are particularly strong in, say, music will appear to be generous in the grading of students in music relative to a different school with the same achievement in mathematics. The same logic applies to schools that are particularly strong in history relative to schools with the same mathematics results but average history performance. Since there are no indicators of performance in the aesthetic and practical subjects, and since the test in the sciences and social sciences are less reliable, there is no real solution to this problem. I will return to this weakness of the approach when discussing the results. However, it should be noted that the three largest school groups – Academedia, IES, and Kunskapsskolan – generally do not offer aesthetic or athletic profiles. In the group of Other free schools, such profiles do exist and there are also some municipal schools with specialized profiles.

5. The data

The data used in this study refers to subject grades and test results during the final year of compulsory schooling in Sweden for the years 2013 to 2016. All data is publicly available through the Siris/SALSA databases maintained by the Swedish National Agency for Education.[11] These databases contain school level information on the average final grades in all subjects and average test grades in the subjects for which national tests are administered, provided that grades or test grades are recorded for at least 10 students. School level data on parental educational attainment, the share of recent immigrant, and the share of boys are also available. The share of test takers are derived by comparing the reported number of test takers to the number of students with grades.[12] The files also contain data on the share of certified teachers in each subject.[13]

The summary statistics for each school category are shown in Table 2. In total, there are 90 180 subject-school-year observations out of which 78 percent refer to public schools. If we were to restrict the sample to observations with test results in the sciences and social sciences, the number drops to just below 62 000 but the relative proportion between school groups is basically unaffected. The table shows clearly that both subject grades and test grades are higher for free schools than for municipal ones, with the IES on top in both regards. The table also shows that a socio-economic index is higher among free schools, again with the IES on top.[14] The share of boys is higher at the municipal schools

[11] Prior to 2013, average grades and test grades were not reported at the school level. All data can be accessed at http://siris.skolverket.se/siris/f?p=SIRIS:108:0::NO ('Till filerna på skolnivå').

[12] The data is cleaned by setting the share of test takers to missing if it is above 112 percent (due to student mobility, the number can exceed 100 percent) or below 40 percent.

[13] For subject-years when data on teacher certification is missing, I impute values using the following procedure. First I use the school level mean value of teacher certification for non-missing years within each subject. For the cases this does not work, I impute the mean school the share of certified teachers for other subjects within the following subject groups: core, science, social science, and aesthetic/practical subjects.

[14] This index is derived by predicting grades using the level of parental educational attainment and the share of newly arrived migrants.

and the share of certified teachers is lower among free schools. It should be noted that the exceptionally low number of certified teachers at the IES to a large extent is due to issues related to the certification of its many teachers of anglo-saxon origin.

Table 2. Summary Statistics

	All	Public	Other	Acade-Media	IES	Kunskaps-skolan
Subject grades	13.70	13.43	14.56	14.37	15.73	15.02
	(1.75)	(1.83)	(1.86)	(1.65)	(1.47)	(1.45)
Test grades	11.90	11.65	12.76	12.40	14.39	12.51
	(2.22)	(2.08)	(2.47)	(2.27)	(1.73)	(2.37)
Share tested	0.92	0.92	0.94	0.93	0.95	0.94
	(0.07)	(0.07)	(0.08)	(0.06)	(0.04)	(0.05)
SES index	13.36	13.12	14.24	13.98	14.43	14.24
	(1.17)	(1.09)	(1.11)	(0.87)	(0.68)	(0.8)
Education level	2.25	2.21	2.40	2.35	2.48	2.40
	(0.23)	(0.21)	(0.24)	(0.2)	(0.12)	(0.18)
Immigrant share	0.04	0.05	0.01	0.01	0.02	0.00
	(0.06)	(0.06)	(0.03)	(0.02)	(0.02)	(0.01)
Share boys	0.51	0.52	0.49	0.50	0.47	0.45
	(0.09)	(0.08)	(0.13)	(0.13)	(0.07)	(0.08)
Share certified	0.60	0.63	0.50	0.45	0.09	0.55
	(0.35)	(0.34)	(0.38)	(0.35)	(0.23)	(0.37)
Observations	90180	70065	14423	2797	1216	1679

Note: The unit of observation is school-by-subject-by-year. Standard deviations in parentheses. The SES index is derived by predicting grades using parental education level and the share of newly arrived migrants. Education level is based on both parents highest educational attainment and runs from 1 to 3. Immigrant share is the share of students who have immigrated less than four years prior to completing compulsory school. Share certified is the share of teachers who are certified in the respective subject.

6. Differences in grading standards

When analyzing differences in grading standards between providers, I test the hypotheses that the four groups of private providers set seminar grades as municipal schools when holding the results on the standardized tests constant. If an estimated coefficient is positive, this means that the provider set higher grades than municipal schools and if the coefficient is negative, the opposite is the case. The main results are presented in graphs and the bars represent the β_j-coefficients from the above equation.[15]

6.1. Core subjects

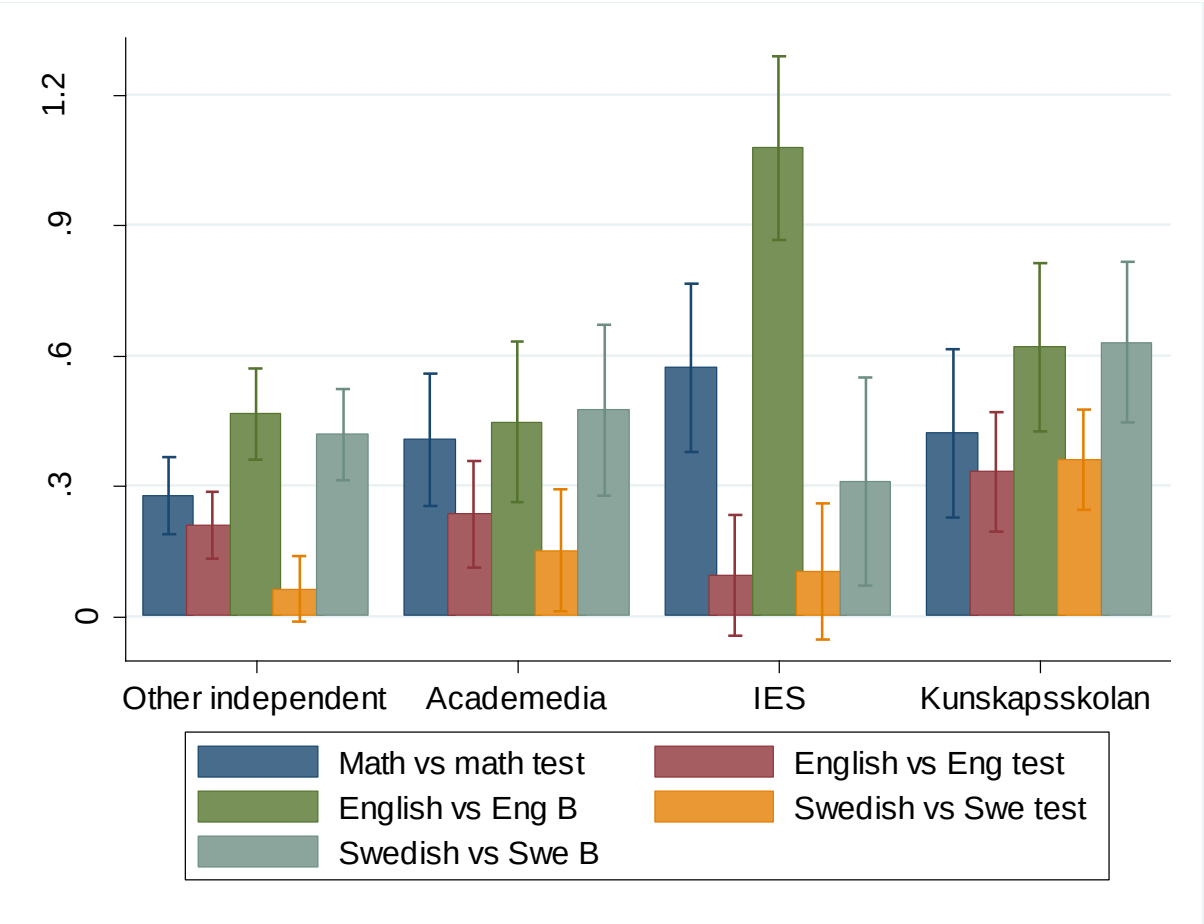
Starting off with the core subjects, I test grade setting relative to the standardized tests. For mathematics there is only one test result available, but for English and Swedish three tests (and the composite full test grade) are available. The most reliable is the B-test of reading and listening comprehension but this test is more limited in scope than the full test. I therefore present results using both the full test and only the reliable B-test.

The results displayed in Figure 1 show that there is a general tendency for all free school groups to set higher grades than municipal schools. In mathematics, the difference ranges from 0.3 grade points

[15] Tables corresponding to each graph can be found in the appendix.

among the group of Other free schools to 0.6 for the IES. This can be compared to a between school standard deviation in math grades of 1.7.

Figure 1. Core subjects



In English, the results depend on whether the full test or B-test is used. For all free school groups, the grading difference relative to municipal schools is significantly larger when the more reliable B-test is used as a comparison. This is particularly stark for the IES which is basically on par with municipal schools when adjusting for the full test score. When using only the B-test, grades among the IES schools are a full 1.1 grade points higher than municipal schools (the standard deviation in English grades is 1.8). Using the full test, the grading differences for the other providers lie between 0.2 and 0.3 grade points. Using the B-test, the difference is between 0.4 (Academedia) and 0.6 (Kunskapsskolan) grade points.

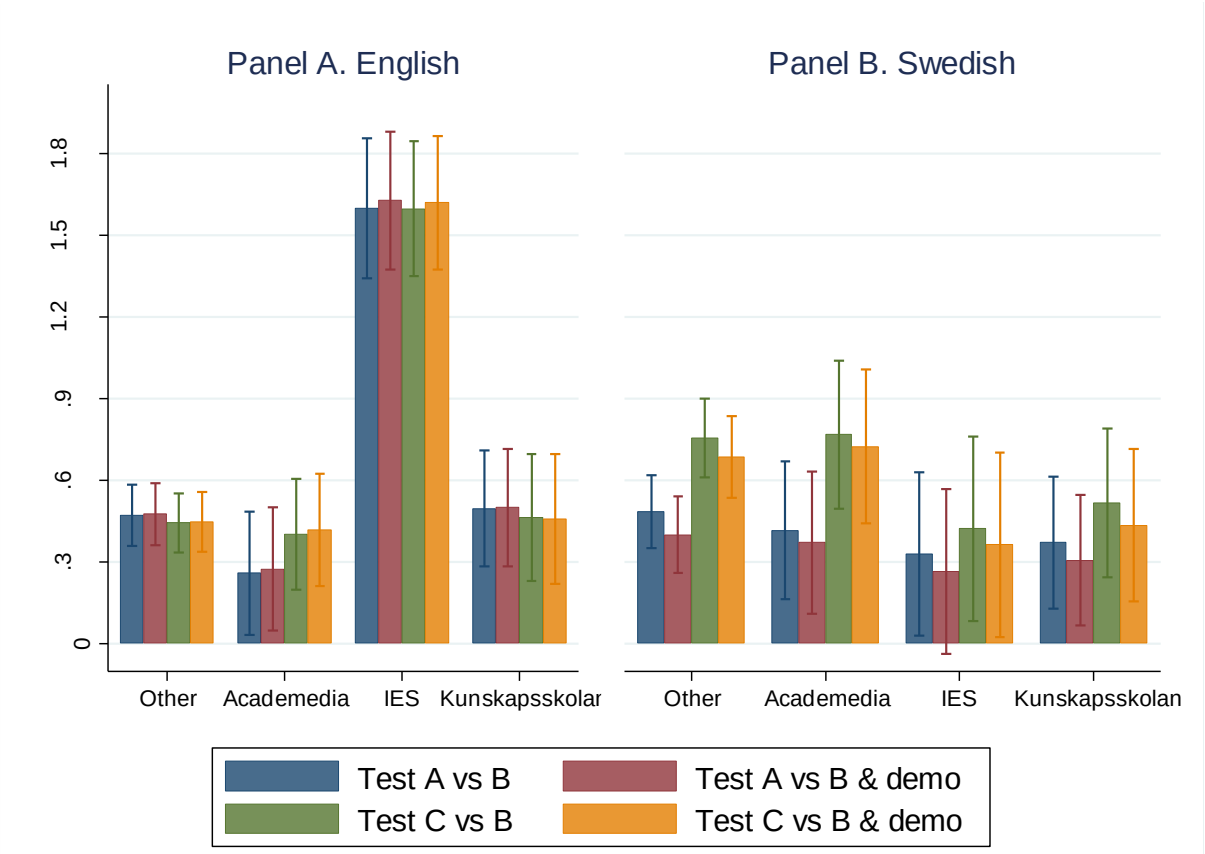
A similar pattern is seen for Swedish. When using the full test, grade setting among both Other free schools and the IES is close to that of municipal schools while the schools belonging to Academedia (0.15) and Kunskapsskolan (0.4) appear more generous. When using the B-test, the difference ranges from 0.3 (IES) to 0.6 (Kunskapsskolan) grade points.

In both English and Swedish the estimated differences between free and municipal schools differ substantially in magnitude when the reliable B-tests are used rather than the less reliable full tests. It is not obvious how these results should be interpreted. In particular, the IES is a bilingual school and it could be that the students at the IES perform relatively well on the verbal and writing tests compared to the test on reading and listening comprehension. Another interpretation is that the grading of the A- and C-tests where teachers have substantial discretion is particularly generous. In principle, it is

impossible to distinguish between more generous test grading and that students at free schools perform particularly well on the less reliable tests.

One reason behind asymmetric performances on the reliable and less reliable tests could be differences in student composition. In Figure 2, I therefore use the same approach as before and regress the results on the less reliable A- and C-tests on a flexible function of results on the more reliable B-tests. I do this both with and without controls for student demographics, i.e. parental education level, the share of newly arrived migrants, and the share of boys. In both English and Swedish, all free school groups score higher on the tests with more teacher discretion compared to municipal schools when adjusting for the more reliable B-tests. These differences are only marginally affected by taking differences in student demographics into account.

Figure 2. English and Swedish A- and C-tests vs B-test

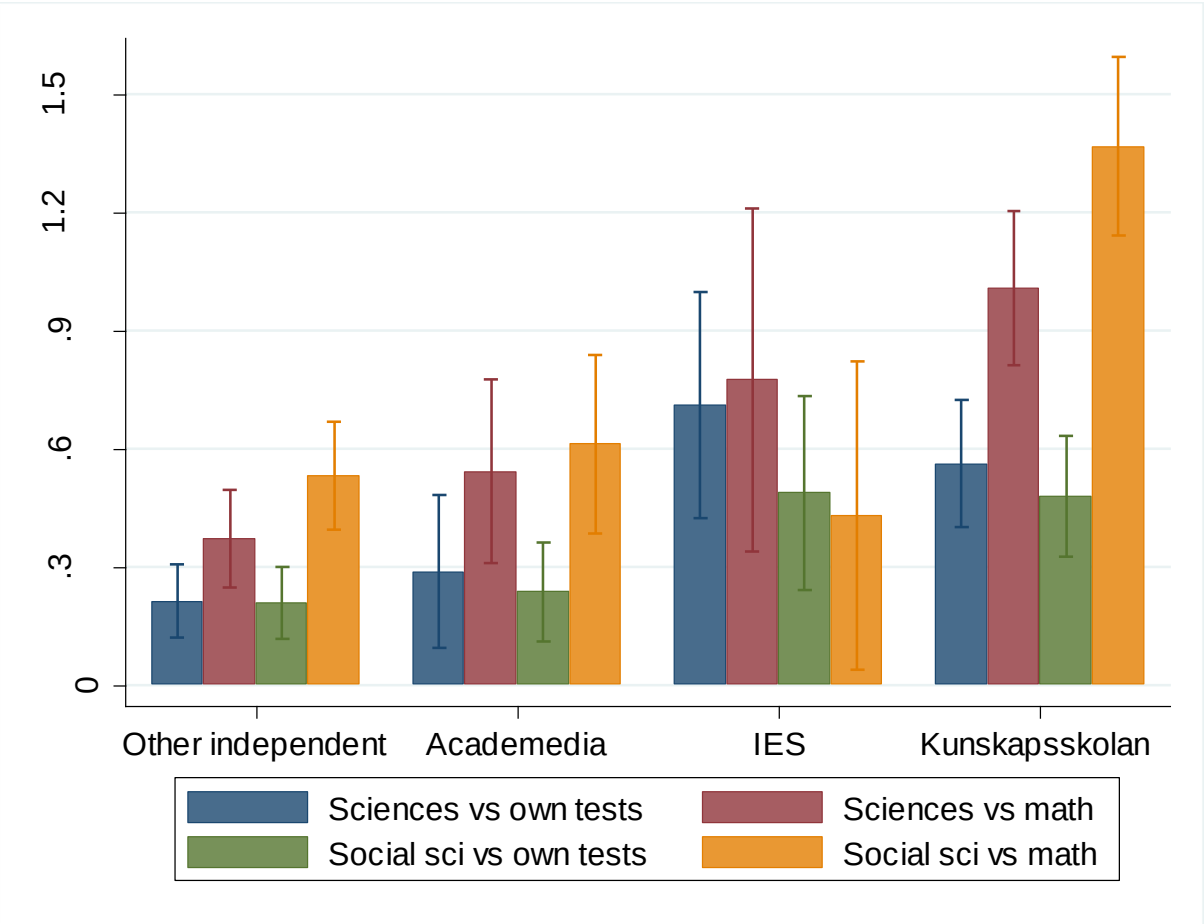


The results in this section clearly indicate that the different groups of free schools award higher grades in the core subjects than municipal when controlling for the results on the standardized tests. The results also show that students at the free school groups have particularly high test results on the tests with more teacher discretion and that these asymmetries cannot be accounted for by differences in student demographics. However, it is plausible that the large differences between the more and less reliable test in English for the IES could at least in part be due to this school being bilingual.

6.2. Grade setting in other subjects

Grade setting is the sciences (biology, chemistry, physics) and the social sciences (civics, geography, history, religion) can be related to the rotating tests in these subjects. One drawback of this approach is that a substantial share of observations is lost. An even more severe problem is that the grading of these tests is highly discretionary and the test results are not very reliable. I therefore relate grades in these subjects first to the smaller sample of own subject tests and next to mathematics test for the full sample of schools. These results are shown in Figure 3.[16]

Figure 3. The sciences and social sciences



Regardless if holding the results on the subject or mathematics tests constant, all free school groupings set higher grades than municipal schools in both the sciences and the social sciences. This is also true when analyzing each of the seven subjects individually: all of the 56 estimates are positive and 49 are statistically significant at the 5-percent level (see the appendix). The magnitudes are substantially larger when using the mathematics test rather than the subject tests as the baseline, however. Test reliability is a serious concern regarding these subject tests. As Hinnerich and Vlachos (2013) have documented, there are substantial differences in how municipal and free schools grade the science tests.[17] As before, it is not possible to distinguish to what extent these differences arise

[16] In the appendix table A4, the results are also shown for each individual subject.

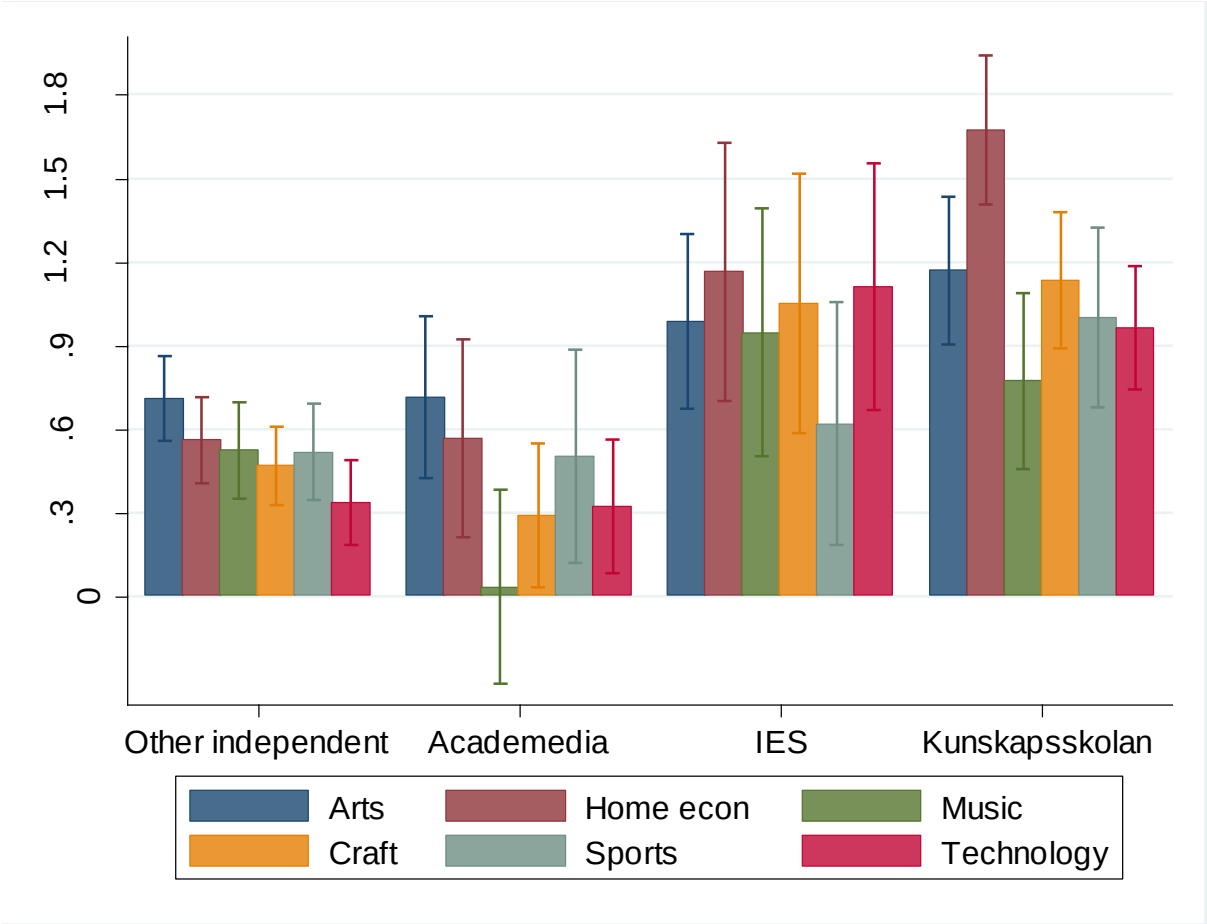
[17] This comparison is done using external and internal evaluations of the exact same tests. No such comparison has been undertaken for the tests in the social sciences.

because students at free schools perform asymmetrically well in these subjects relative to their performance in mathematics or because of different grading standards.

Results for the six aesthetic and practical subjects (arts, home economics, music, craft, sports, technology) are displayed in Figure 4. Since there are no standardized tests in these subjects, differences in grading standards are estimated by controlling for the test results in mathematics. This is the most reliable test and available for all school-years but it is obviously not a perfect measure of subject-specific achievement in aesthetic and practical subjects. It should therefore be stressed that there may be perfectly legitimate for some schools to have high grades in these subjects relative to their achievement in mathematics.

This caveat in mind, the results are striking. With the exception of music grades among the Academedia schools, all estimates are positive and statistically significant. Many estimates are also sizeable, in particular for the IES and Kunskapsskolan. For the IES, the estimates range from 0.6 (sports) to 1.2 (home economics) grade points relative to municipal schools. For Kunskapsskolan, the range in from 0.8 (music) to 1.7 (home economics) grade points. This can be compared to a standard deviation in grades for these subjects which is around 1.6 grade points.

Figure 4. Aesthetic and practical subjects



One potential explanation for differences in grading between free and municipal schools for these subjects is that a larger share of free schools specializes in aesthetic subjects and sports. That students in such specialized programs perform relatively well in these subjects relative to their performance in mathematics would be unsurprising. However, the free schools with such profiles are overwhelmingly in the Other group. Given the profiles of the large corporate groups, there is no *ex ante* reason to

expect their students to be particularly outstanding in these subjects relative to their performance in mathematics.

6.3. Accounting for differences in grading standards

Across essentially all subjects, all four groups of free schools tend to set higher grades than municipal schools when controlling for academic achievement and these differences are consistently larger when the more reliable tests are used as controls. In this section, I analyze to what extent these differences can be accounted for by the location and student composition of the free schools. First, free schools are highly concentrated to relatively highly populated areas. In such regions, competition between schools is stronger and students' face more choices at the upper-secondary level. Since admittance to upper-secondary schools is based on compulsory school GPA, students' incentives to perform well across all subjects may differ. Second, student demographics differ quite substantially between municipal and free schools. Such differences in demographics may also be related to differences in the relation between subject and test grades.[18] Third, teacher certification is lower among free schools and this can potentially affect grade setting.

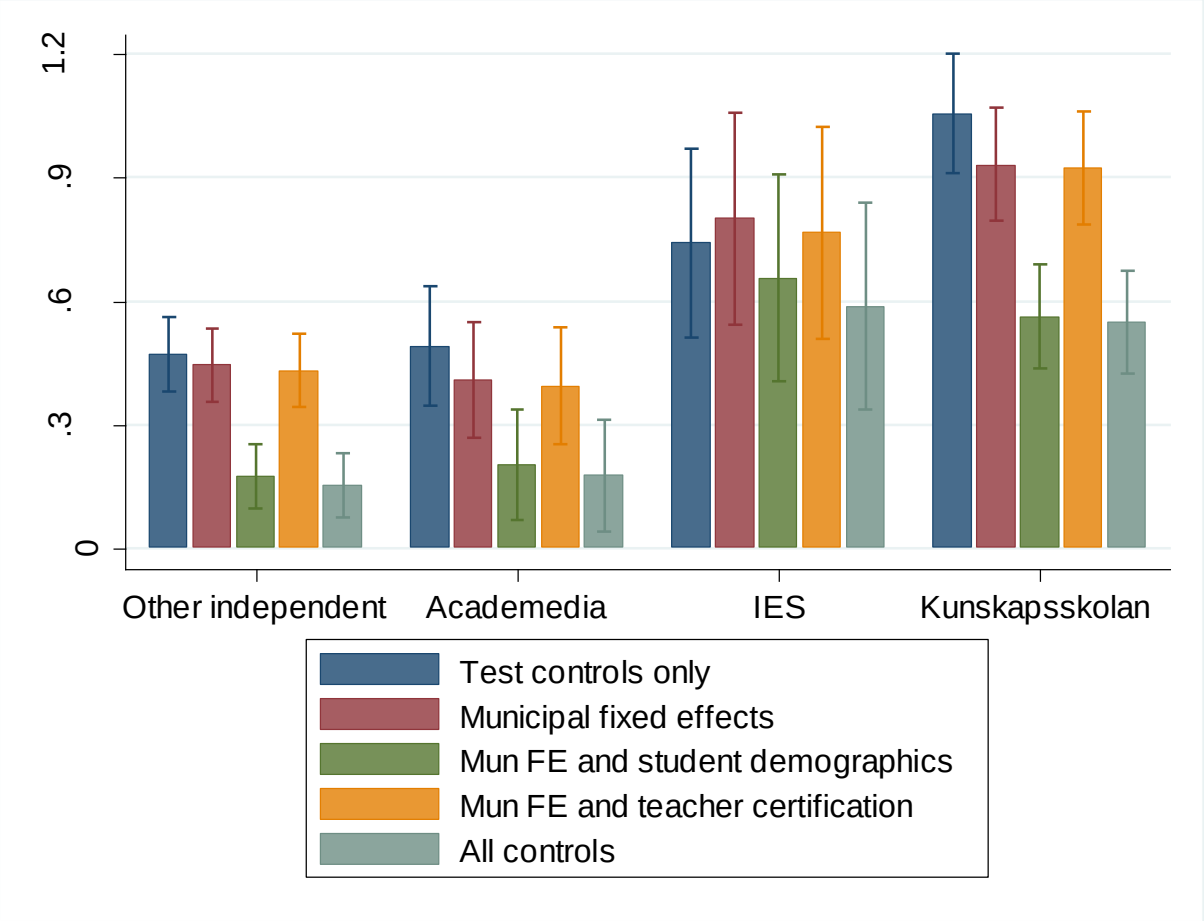
It should be clear that this is an accounting exercise and not an attempt to isolate the causal impact of various school characteristics on grade setting. For example, schools decide which fraction of certified teachers to employ. Should this variable account for grading differences between providers, this just highlights one of the possible mechanisms through which schools can influence grading standards. Similarly, students with certain characteristics may self-select to schools depending on the schools' grading standards. This reasoning also applies to location: if free schools are located in regions where grading is more lenient, this could in part be due to competition induced by these schools. Hence, it is not possible to separate the impact of schools from the characteristics of location, teachers and students.

The results pooled across all subjects are shown in Figure 5. The first set of bars shows the estimates when only test results are controlled for.[19] Next, adding municipal fixed effects reduces the estimates somewhat for all free school groups except the IES. This indicates that free schools on average tend to be located in high-grading areas. As shown by the third set of bars, adding controls for student demographics is more important. All point estimates are reduced and among several school groups quite substantially. All estimates are still positive and statistically significant, however, and they are still large for the IES and Kunskapsskolan (0.6-0.7 grade points) and around 0.2 grade points for the other two groups of free schools. The final two sets of bars show that controlling for teacher certification only has a marginal impact on the estimates.

[18] For example, it well known that girls tend to receive higher grades relative to test results than boys (Voyer and Voyer, 2014). See also Table A10 in the appendix.

[19] I here use the most reliable tests, i.e. the B-tests in English and Swedish and the mathematics test for the sciences, social sciences, and aesthetic/practical subjects. For other specifications, see table A6.

Figure 5. Accounting for differences in grading standards (pooled)



This analysis is also conducted separately for each of the four subject groups (core, sciences, social sciences, aesthetic/practical). Focusing on the specifications with municipal fixed effects and demographic controls, the results are similar to the above. For core subjects, all estimates are positive and statistically significant ranging from 0.2 (Other) to 0.6 (IES). In the sciences, the estimate is close to zero (0) for the Other group, while it ranges from 0.2 (Academedia) to 0.7 (IES) for the other groups of providers. In the social sciences, the estimates fall between is 0.2 (Other) and 0.8 (Kunskapsskolan). For the practical subjects, the estimate is close to zero (0) for Academedia, 0.2 for Other, 0.6 for Kunskapsskolan, and 0.9 for the IES.[20] As mentioned before, there is nothing in the profiles of the IES and Kunskapsskolan to suggest that their students should be particularly accomplished in the aesthetic or practical subjects.

The summary statistics in Table 2 reveal that subject grades are higher among free schools than among municipal schools. Students at free schools are also positively selected with respect to socio-economic and migrant background. When adjusting subject grades to location and socio-economic characteristics, free schools that belong to the Academedia and Other groups outperform municipal schools by 0.2 grade points (Table A6, column 9). This is close to identical to the estimated grading leniency reported above for these groups. Using the same approach, Kunskapsskolan outperforms

[20] Again, I refer the reader to the appendix for details (Table A7). Note that the estimate of 0.3 for the IES in the social sciences is not statistically significant. The appendix also shows estimates with municipal fixed effects and demographic controls for all 16 subjects (Table A8). Out of 64 estimates, 44 are positive and statistically significant, 19 are not statistically significant (none of these is negative), and one is negative and statistically significant (music for Academedia).

municipal schools by 0.5 grade points, while the above estimates suggest a grading leniency of 0.6 grade points. For schools that belong to the IES, the advantage over municipal schools amounts to 1.1 grade points which can be related to an estimated grading leniency of 0.7 grade points. Among all free school groups except the IES, the entire advantage over municipal schools can thus be attributed to greater leniency in grading. For the IES, almost two-thirds of the advantage over municipal schools can be attributed more generous grading.

7. Evidence from upper-secondary schools

Free schools have a larger market share at the upper-secondary level and the student selection differs markedly from the compulsory level. While the socio-economic selection of students to free schools tends to be positive at the compulsory level, the selection is negative at the upper-secondary level both with respect to socio-economic background and prior levels of academic achievement (Hinnerich and Vlachos, 2017). The analysis in this chapter is restricted to grading standards at the compulsory level but some previous studies have analyzed related questions at the upper-secondary level.

Wikström and Wikström (2005) find that students' from upper-secondary academic programs at free schools tend to have substantially higher GPA compared to their results on an SAT-equivalent test (Swe-SAT). While indicative of greater leniency in grading, the Swe-SAT captures different types of abilities and subject knowledge than the GPA (Cliffordson, 2008). Other issues are that students self-select into taking the Swe-SAT, and that just three percent of students attended upper-secondary free schools in the year (1997) that Wikström and Wikström study. In a more recent study, Hinnerich and Vlachos (2017) compare the results of internal and external evaluations of the exact same standardized tests. The results show that free schools tend to grade the tests more leniently than municipal schools, also when controlling for a host of student level characteristics and prior achievement.

Both studies thus suggest lower grading standards among upper-secondary free schools. These results are perfectly compatible with a recent government report (Skolverket, 2018) showing that students from academic programs at free schools tend to underperform during their first year of tertiary education compared to students from municipal schools. In the report, upper-secondary GPA and a number of demographic and socio-economic indicators are controlled for.

8. Conclusion

The Swedish school system has somewhat oddly combined market principles such as decentralization, choice, competition, and corporate providers with an evaluation system that is highly trust-based and where teacher-set school grades are high-stakes for the students. This means that both students and schools have incentives to game a system that is easy to game and the findings suggest that the integrity of the evaluation system has been compromised. The results show that all groupings of free schools set higher grades than municipal schools when controlling for student achievement on national tests. As the national tests are locally graded, they are not fully reliable and the differences between public and private providers are more pronounced when more reliable tests are used to control for achievement.

To some extent, the differences in grading standards between municipal and free schools can be accounted for by differences in location and student demographics. Even after holding such factors constant, however, grading standards among private providers appear lenient, in particular among schools that belong to two of the large corporate groups (IES and Kunskapsskolan). Grading is less lenient among other free schools, including those that belong to the large corporate group

Academedia, but still significantly more generous than among municipal schools. These differences between corporate providers that face similar incentives seem worthy of deeper investigation.

Students at all free schools groupings have higher teacher-set subject grades than students at municipal schools, an advantage that remains after taking location and student composition into account. Among all free school groupings except the IES, however, the results indicate that this performance advantage over public schools can fully be accounted for by more lenient grading standards. For free schools that belong to the IES, close to two-thirds of the grading advantage over public schools can be attributed to more generous grading.

Some caveats are worth mentioning. First of all, no test is perfect and there is a trade-off between the reliability and validity of the tests. This is to say that a test with little scope of teacher discretion in grading most likely does not capture the full scope of what is ideally to be tested. In much of the analysis, I have relied on the most reliable tests when estimating differences in grading standards. It could be that some of the differences between free and municipal schools are due to this choice, but there is no way of knowing the direction of the bias that using the more reliable tests would result in. Second, even the more reliable tests locally graded and some scope for discretion is always present. The results from other studies indicate that free schools are on average more lenient when grading the tests. This suggests that the difference in grading standards between private and municipal providers is actually underestimated, but this remains an open question. Finally, there may be systematic differences in how different providers engage in exaggerated teaching-to-the-tests or excessive assistance during the actual tests. Such differences are not captured by this study and, again, the direction of the biases stemming from such factors is unknown.

Back in the 1970's, Campbell (1979, p. 85) wrote that 'the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the processes it is intended to monitor'. Campbell's argument was not only that certain incentives would corrupt the measurement of – for example – educational outcomes, but also that education itself might be harmed. In this chapter, I have presented evidence suggesting that opening up education to providers with strong incentives can compromise educational measurement. Whether it also corrupts the educational process itself is an even more important question, but a question I leave for others to consider. An important takeaway from the analysis presented here is that different providers do not necessarily respond in the same way, even when faced with similar market conditions and acting under the same regulatory regime.

References

Abrams, Samuel (2016). Education and the Commercial Mindset. Harvard University Press.

Campbell, Donald T. (1979). Assessing the Impact of Planned Social Change. Evaluation and Program Planning, 2(1), 67-90.

Cliffordson, Christina (2008). Differential prediction of study success across academic programs in the swedish context: the validity of grades and tests as selection instruments for higher education. Educational Assessment, 13:1, 56-75.

Friedman, Milton (1962). Capitalism and Freedom, University of Chicago Press.

Friskolornas Riksförbund (2017). Fakta om Friskolor, mars 2017.

Hoxby, Caroline (2003). School choice and school productivity. Could school choice be the tide that lifts all boats in Hoxby (ed.) The Economics of School Choice, University of Chicago Press.

Jacob, Brian, Brian McCall, and Kevin Stange (2018). College as a Country Club: Do Colleges Cater to Students' Preferences for Consumption? Journal of Labor Economics, 36(2), 309-348.

Koretz, Daniel (2017). *The Testing Charade. Pretending to Make Schools Better*. University of Chicago Press.

Skolinspektionen (2013a). Beslut om ansökningar att starta fristående skola läsåret 2013/14. Dnr 30-2013:1176.

Skolinspektionen (2013b). Olikheterna är för stora: Omrättning av nationella prov i grund- och gymnasieskola, 2013. Redovisning av regeringsuppdrag Dnr U2011/6544/GV.

Skolverket (2018), "Från gymnasieskola till högskola – en registerstudie", Rapport 466.

Hinnerich Tyrefors, Björn and Jonas Vlachos (2013). Systematiska skillnader mellan interna och externa bedömningar av nationella prov – en uppföljningsrapport. Appendix 5 to Skolinspektionen (2013b).

Hinnerich Tyrefors, Björn and Jonas Vlachos (2017). The Impact of Upper-Secondary Voucher School Attendance on Student Achievement. Swedish Evidence Using External and Internal Evaluations. Labour Economics, 47, 1-14.

Shleifer, Andrei (1998). State versus Private Ownership. Journal of Economic Perspectives, 12:4, 133-150.

Voyer, Daniel and Susan Voyer (2014). Gender Differences in Scholastic Achievement: A Meta-Analysis. Psychological Bulletin, 140:4, 1174-1204.

Appendix (for working paper version)

Table A1. Core subjects

VARIABLES	(1) Mathematics	(2) English	(3)	(4) Swedish	(5)
Other free	0.28**	0.21**	0.47**	0.06	0.42**
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Academedia	0.41**	0.23**	0.45**	0.15*	0.47**
	(0.08)	(0.06)	(0.09)	(0.07)	(0.10)
IES	0.57**	0.09	1.08**	0.10	0.31*
	(0.10)	(0.07)	(0.11)	(0.08)	(0.12)
Kunskapsskolan	0.42**	0.33**	0.62**	0.36**	0.63**
	(0.10)	(0.07)	(0.10)	(0.06)	(0.09)
Observations	5,648	5,648	5,648	5,472	5,472
R-squared	0.76	0.86	0.76	0.77	0.60
Test controls	Math	Eng	Eng B	Swe	Swe B

Note: The dependent variable is subject grades. Fixed subject-by-year effects included and interacted with the respective test result and the share of test takers. Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A2. English and Swedish A- and C-tests vs B-test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	English A		English C		Swedish A		Swedish C	
Other free	0.47**	0.48**	0.44**	0.45**	0.48**	0.40**	0.75**	0.68**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)
Academedia	0.26*	0.27*	0.40**	0.42**	0.42**	0.37**	0.77**	0.72**
	(0.12)	(0.12)	(0.10)	(0.11)	(0.13)	(0.13)	(0.14)	(0.14)
IES	1.60**	1.63**	1.60**	1.62**	0.33*	0.26	0.42*	0.36*
	(0.13)	(0.13)	(0.13)	(0.12)	(0.15)	(0.15)	(0.17)	(0.17)
Kunskapsskolan	0.50**	0.50**	0.46**	0.46**	0.37**	0.31*	0.52**	0.43**
	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.14)	(0.14)
Observations	5,648	5,648	5,648	5,648	5,472	5,472	5,472	5,472
R-squared	0.65	0.65	0.70	0.70	0.43	0.44	0.47	0.47
Test controls	Eng B	Eng B	Eng B	Eng B	Swe B	Swe B	Swe B	Swe B
Demographics	No	Yes	No	Yes	No	Yes	No	Yes

Note: The dependent variable is the A- or the C-test result in the respective subject. Fixed subject-by-year effects included and interacted with test controls (the B-test result in the respective subject). Demographic variables are parental educational level, the share of recent immigrants, and the share of boys (interacted with the fixed subject-by-year effects where applicable). Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A3. Sciences and social sciences (by group)

	(1)	(2)	(3)	(4)
VARIABLES	Sciences		Social sciences	
Other free	0.21**	0.37**	0.21**	0.53**
	(0.05)	(0.06)	(0.05)	(0.07)
Academedia	0.29**	0.54**	0.24**	0.61**
	(0.10)	(0.12)	(0.06)	(0.12)
IES	0.71**	0.78**	0.49**	0.43*
	(0.15)	(0.22)	(0.13)	(0.20)
Kunskapsskolan	0.56**	1.01**	0.48**	1.37**
	(0.08)	(0.10)	(0.08)	(0.12)
Observations	5,624	16,944	5,621	22,592
R-squared	0.75	0.53	0.73	0.48
Test controls	Sciences	Math	Social sciences	Math

Note: The dependent variable is subject grades. Fixed subject-by-year effects included and interacted with test controls and the share of test takers. Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A4. Sciences and social sciences (by subject)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Biology		Chemistry		Physics		Civics		Geography		History		Religion	
Other free	0.32**	0.41**	0.16*	0.38**	0.16*	0.33**	0.34**	0.57**	0.19*	0.47**	0.15	0.55**	0.17*	0.53**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)
Academedia	0.28*	0.52**	0.45**	0.61**	0.15	0.50**	0.25*	0.58**	0.33**	0.65**	0.27	0.63**	0.10	0.59**
	(0.11)	(0.12)	(0.15)	(0.14)	(0.13)	(0.12)	(0.11)	(0.11)	(0.11)	(0.13)	(0.16)	(0.12)	(0.13)	(0.13)
IES	0.67**	0.69**	0.64**	0.84**	0.81**	0.80**	0.52*	0.45*	0.50**	0.44*	0.39	0.37	0.56*	0.46*
	(0.22)	(0.22)	(0.18)	(0.22)	(0.19)	(0.23)	(0.22)	(0.20)	(0.13)	(0.22)	(0.23)	(0.19)	(0.26)	(0.21)
Kunskapsskolan	0.54**	1.01**	0.57**	1.05**	0.58**	0.96**	0.65**	1.46**	0.42**	1.51**	0.62**	1.27**	0.24	1.23**
	(0.09)	(0.11)	(0.11)	(0.12)	(0.16)	(0.10)	(0.13)	(0.13)	(0.15)	(0.12)	(0.15)	(0.13)	(0.14)	(0.12)
Observations	1,900	5,648	1,857	5,648	1,867	5,648	1,391	5,648	1,442	5,648	1,389	5,648	1,399	5,648
R-squared	0.73	0.52	0.76	0.53	0.75	0.54	0.77	0.49	0.76	0.48	0.65	0.47	0.74	0.46
Test controls	Biol	Math	Chem	Math	Physics	Math	Civics	Math	Geo	Math	Hist	Math	Rel	Math

Note: The dependent variable is subject grades. Fixed subject-by-year effects included and interacted with test controls and the share of test takers. Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A5. Aesthetic and practical subjects

VARIABLES	(1) Arts	(2) Home Econ	(3) Music	(4) Craft	(5) Sports	(6) Technology
Other free	0.71**	0.56**	0.53**	0.47**	0.52**	0.34**
	(0.08)	(0.08)	(0.09)	(0.07)	(0.09)	(0.08)
Academedia	0.72**	0.57**	0.03	0.29*	0.50**	0.32**
	(0.15)	(0.18)	(0.18)	(0.13)	(0.19)	(0.12)
IES	0.99**	1.16**	0.95**	1.05**	0.62**	1.11**
	(0.16)	(0.24)	(0.23)	(0.24)	(0.22)	(0.23)
Kunskapsskolan	1.17**	1.67**	0.77**	1.13**	1.00**	0.96**
	(0.13)	(0.14)	(0.16)	(0.12)	(0.16)	(0.11)
Observations	5,646	5,647	5,647	5,643	5,646	5,647
R-squared	0.31	0.34	0.28	0.32	0.38	0.40
Test controls	Math	Math	Math	Math	Math	Math

Note: The dependent variable is subject grades. Fixed subject-by-year effects included and interacted with test controls and the share of test takers. Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A6. Accounting for grading differences (pooled)

VARIABLES	(1) Pooled	(2) Pooled	(3) Pooled	(4) Pooled	(5) Pooled	(6) Pooled	(7) Pooled	(8) Pooled	(9) Pooled
Other free	0.47**	0.44**	0.17**	0.43**	0.15**	0.15**	0.14**	0.18**	0.19**
	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)
Academedia	0.49**	0.41**	0.20**	0.39**	0.18*	0.18**	0.11	0.14*	0.18*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.09)
IES	0.74**	0.80**	0.66**	0.77**	0.59**	0.59**	0.64**	0.73**	1.12**
	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.09)	(0.10)	(0.13)
Kunskapsskolan	1.06**	0.93**	0.56**	0.92**	0.55**	0.54**	0.41**	0.45**	0.48**
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Observations	90,180	90,180	90,180	90,180	90,180	90,180	61,889	61,889	90,180
R-squared	0.55	0.58	0.62	0.58	0.62	0.63	0.66	0.64	0.56
Test controls	Core B	Core B	Core B	Core B	Core B	Core	Own	Own B	None
Muncipal effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student demographics	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Teacher certification	No	No	No	Yes	Yes	No	No	No	No

Note: The dependent variable is subject grades. Test controls in (1)-(5) are the mathematics test for mathematics, the sciences, the social sciences, and aesthetic/practical subjects. The share of test takers is controlled for. For English and Swedish, the respective B-test is used. In (6), the full subject tests in English and Swedish are used. In (7) the full tests are used for core subjects and the subject tests in the sciences and social sciences. In (8), the B-tests in English and Swedish are used and the science and social science tests for the respective subject. In (9), test results are not controlled for. Demographic variables are parental educational level, the share of recent immigrants, and the share of boys. subject-by-year effects included (interacted with the control variables where applicable). Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A7. Accounting for grading differences (by subject group)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
VARIABLES	Core				Sciences				Social sciences				Aesthetic/Practical			
Other free	0.38**	0.34**	0.18**	0.17**	0.37**	0.35**	0.04	0.01	0.53**	0.51**	0.17**	0.17**	0.52**	0.49**	0.24**	0.21**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
Academedia	0.44**	0.39**	0.29**	0.26**	0.54**	0.49**	0.24*	0.20	0.61**	0.55**	0.30*	0.30*	0.41**	0.28**	0.08	0.04
	(0.06)	(0.07)	(0.07)	(0.07)	(0.12)	(0.12)	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.10)	(0.08)	(0.08)	(0.08)
IES	0.64**	0.67**	0.61**	0.51**	0.78**	0.85**	0.68**	0.55*	0.43*	0.46*	0.28	0.27	0.98**	1.05**	0.91**	0.83**
	(0.08)	(0.09)	(0.09)	(0.09)	(0.22)	(0.22)	(0.22)	(0.22)	(0.20)	(0.22)	(0.21)	(0.21)	(0.11)	(0.12)	(0.12)	(0.12)
Kunskapsskolan	0.55**	0.48**	0.29**	0.29**	1.01**	0.92**	0.49**	0.47**	1.37**	1.26**	0.80**	0.80**	1.12**	0.93**	0.57**	0.55**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)	(0.10)	(0.12)	(0.12)	(0.11)	(0.11)	(0.09)	(0.09)	(0.09)	(0.09)
Observations	16,768	16,768	16,768	16,768	16,944	16,944	16,944	16,944	22,592	22,592	22,592	22,592	33,876	33,876	33,876	33,876
R-squared	0.78	0.80	0.82	0.82	0.53	0.58	0.63	0.63	0.48	0.54	0.60	0.61	0.37	0.42	0.46	0.46
Test controls	Core B	Core B	Core B	Core B	Math	Math	Math	Math	Math	Math	Math	Math	Math	Math	Math	Math
Municipal effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Teacher certification	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes

Note: The dependent variable is subject grades. Test controls are the mathematics test for mathematics, the sciences, the social sciences, and aesthetic/practical subjects. The share of test takers is controlled for. For English and Swedish, the respective B-test is used. Demographic variables are parental educational level, the share of recent immigrants, and the share of boys. Fixed subject-by-year effects included (interacted with the control variables where applicable). Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A8. Demographic controls and municipal fixed effects (by subject)

VARIABLES	(1) Mathematics	(2) English	(3) Swedish	(4) Biology	(5) Chemistry	(6) Physics	(7) Civics	(8) Geography
Other free	0.00	0.33**	0.24**	0.05	0.05	0.00	0.22**	0.11
	(0.04)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Academedia	0.18*	0.32**	0.36**	0.19	0.31*	0.22	0.26*	0.34*
	(0.09)	(0.09)	(0.11)	(0.12)	(0.13)	(0.12)	(0.12)	(0.13)
IES	0.46**	1.25**	0.22	0.58**	0.73**	0.73**	0.29	0.32
	(0.11)	(0.13)	(0.13)	(0.22)	(0.23)	(0.24)	(0.21)	(0.24)
Kunskapsskolan	0.02	0.43**	0.44**	0.44**	0.52**	0.50**	0.88**	0.95**
	(0.09)	(0.09)	(0.10)	(0.12)	(0.11)	(0.10)	(0.13)	(0.12)
Observations	5,648	5,648	5,472	5,648	5,648	5,648	5,648	5,648
R-squared	0.82	0.80	0.68	0.63	0.63	0.64	0.62	0.62
	(9) History	(10) Religion	(11) Arts	(12) Home Econ	(13) Music	(14) Craft	(15) Sports	(16) Technology
Other free	0.16*	0.19**	0.41**	0.25**	0.19*	0.28**	0.27**	0.05
	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	(0.07)	(0.09)	(0.08)
Academedia	0.29*	0.30*	0.38*	0.27	-0.37*	0.07	0.13	0.00
	(0.12)	(0.13)	(0.15)	(0.17)	(0.16)	(0.13)	(0.19)	(0.13)
IES	0.22	0.30	0.90**	1.07**	0.80**	1.00**	0.58*	1.09**
	(0.20)	(0.22)	(0.21)	(0.24)	(0.25)	(0.24)	(0.24)	(0.22)
Kunskapsskolan	0.69**	0.67**	0.60**	1.14**	0.01	0.70**	0.54**	0.46**
	(0.12)	(0.12)	(0.14)	(0.14)	(0.15)	(0.13)	(0.17)	(0.11)
Observations	5,648	5,648	5,646	5,647	5,647	5,643	5,646	5,647
R-squared	0.61	0.58	0.48	0.49	0.46	0.46	0.53	0.52

Note: The dependent variable is subject grades. Municipal fixed effects and demographic controls (parental educational level, share recent immigrants, share of boys) included. Test controls are the mathematics test for mathematics, the sciences, the social sciences, and aesthetic/practical subjects. For English and Swedish, the respective B-test is used. Share of test takers included as a control variable. Fixed year effects included and interacted with all control variables. Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A9. Sensitivity to control function

	(1)	(2)	(3)	(4)	(5)	(5)
VARIABLES	All subjects pooled					
Other free	0.54**	0.48**	0.47**	0.47**	0.47**	0.47**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Academedia	0.52**	0.49**	0.49**	0.49**	0.50**	0.49**
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
IES	0.90**	0.77**	0.76**	0.74**	0.74**	0.74**
	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Kunskapsskolan	1.09**	1.06**	1.06**	1.06**	1.06**	1.06**
	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)
Observations	90,180	90,180	90,180	90,180	90,180	90,180
R-squared	0.55	0.55	0.55	0.55	0.55	0.56
Specification	Linear	Quadratic	Cubic	Baseline	Quintile	Decile

Note: The dependent variable is subject grades. Test controls are the mathematics test for mathematics, the sciences, the social sciences, and aesthetic/practical subjects. For English and Swedish, the respective B-test is used. Share of test takers included as a control variable. Fixed subject-by-year effects included and interacted with test controls and share of test takers. *Specification* refers to the functional form of the control function. *Baseline* indicates a fourth degree polynomial in test scores, *Quintiles* are quintile group indicators of test scores interacted with average test scores, *Deciles* decile group indicators of test scores interacted with average test scores. Standard errors clustered by school. ** p<0.01, * p<0.05.

Table A10. Demographics and teacher certification

VARIABLES	(1) Pooled	(2) Core	(3) Science	(4) Social	(5) Practical
Education level	1.81**	1.19**	1.94**	2.43**	1.61**
	(0.08)	(0.07)	(0.12)	(0.13)	(0.10)
Share immigrant students	-2.43**	-1.33**	-2.72**	-2.88**	-2.47**
	(0.29)	(0.25)	(0.46)	(0.42)	(0.32)
Share boys	-1.10**	-0.66**	-1.07**	-1.22**	-1.25**
	(0.12)	(0.11)	(0.17)	(0.18)	(0.15)
Share certified teachers	-0.21**	-0.28**	-0.23**	-0.07	-0.26**
	(0.03)	(0.04)	(0.05)	(0.05)	(0.03)
Observations	90,180	16,768	16,944	22,592	33,876
R-squared	0.58	0.79	0.57	0.54	0.39

Note: The dependent variable is subject grades. Test controls are the mathematics test for mathematics, the sciences, the social sciences, and aesthetic/practical subjects. For English and Swedish, the respective B-test is used. Share of test takers included as a control variable. Fixed subject-by-year effects included and interacted with test controls and share of test takers. Standard errors clustered by school. ** p<0.01, * p<0.05