CESifo
AREA CONFERENCES 2021

Economics of Education
Munich, 3–4 September 2021

School value-added and long-term student outcomes

Lars J. Kirkebøen

CESifo

School value-added and long-term student outcomes

Lars J. Kirkebøen*

Abstract

This paper studies school quality in the context of Norwegian compulsory schooling. I demonstrate that even when lagged achievement is not observed, it is possible to construct informative value-added (VA) indicators of persistent school quality by adjusting exam scores for students' background characteristics. These VA indicators show little bias forecasting average exam performance out of sample, and are also predictive of long-term student outcomes, including earnings. Three quasi-experiments using variation from student mobility and changes in neighborhood school assignments indicate that the differences captured by the VA indicators do indeed reflect differences in school quality, rather than unobserved student characteristics. The findings help connect learning outcomes with later labor market outcomes, e.g. for cost-benefit analysis of interventions in schools.

Keywords: School quality, value-added, VAM, earnings

JEL codes: J24, I2

*Statistics Norway, Research Department, e-mail: kir@ssb.no
I thank participants in several seminars for valuable feedback and suggestions. The Norwegian Research Council supported this research under projects no. 238050 and 275906.

1 Introduction

Primary and secondary schooling has in several recent studies shown a great potential for improving long-term student outcomes (Chetty et al. [2011], Fredriksson et al. [2012], Deming et al. [2014], Chetty et al. [2014b]). This has further spurred the interest in interventions that improve students' performance (Roland G. Fryer [2017]) and in identifying effective teachers and schools. Researcher and policy makers alike have long shown interest in the potential of value-added models (VAM) to identify teachers and schools of high quality. In a study of teacher value-added (VA), Chetty et al. [2014a,b] find that VAM controlling for lagged test scores exhibit little bias when used to forecast teacher quality and also that VAM successfully predict long-term outcomes such as college enrollment and earnings. Following the influential study of Kane and Staiger [2008], several studies have investigated school VA. Deming et al. [2014] use random variation in school assignment from school lotteries and find that indicators controlling for lagged scores are informative about school quality, and that changes in school quality have considerable impacts on long-term student outcomes. Angrist et al. [2017, 2020] further develop how school assignment lotteries can be used to provide credible estimate of schools' causal effects on students' outcomes.

In this paper I study school quality in Norwegian compulsory schooling. I construct indicators for persistent VA as leave-out-year indicators, similar to Chetty et al. [2014b], based on test scores, exam scores and teacher grades. I also construct indicators for transitory VA, unexplained performance over and above the persistent VA. Like Chetty et al. [2014b] I investigate whether the indicators forecast school performance out-of-sample. I also study whether the indicators predict longer-term student outcomes, including labor market outcomes. Finally, I study three different quasi-experiments, where students move or are assigned to a different school, to investigate whether the indicators capture causal effects of the schools.

While exam scores and teacher grades from the end of compulsory schooling are available since the early 2000's, standardized tests throughout compulsory school were introduced only from 2007. Thus, controlling for lagged test scores is possible only for relatively recent cohorts, and for these cohorts there is limited information on longer-term outcomes. I therefore construct indicators of school quality by controlling for data on family background only, and compare the performance of these indicators to indicators obtained by also controlling for lagged test scores.

I find that VA indicators controlling either for socio-demographic characteristics or for previous test scores forecast exam scores out-of-sample without bias, conditional on student characteristics. Studying VA at different stages, lagged achievement can be interchanged with the corresponding VA measures without impacting the estimates for later VA. VA at all stages during compulsory schools is predictive for exam scores and high school completion, but VA towards the end of compulsory school more so.

Furthermore, students attending schools with higher estimated persistent VA based on exam scores have better progress through upper secondary school, are more likely to complete upper secondary school, are less likely to be observed as neither employed nor under education in early adulthood, and obtain higher earnings in the labor market. The correlations between outcomes and VA indicators are large relative to the corresponding cross-sectional student-level correlations between outcomes and exam scores. On the other hand, the correlations between VA based on teacher grades and outcomes other than teacher grades are much lower than the corresponding correlations with outcomes based on exam scores, despite the individual-level correlations with teacher grades being higher.

In the final part of the paper I draw on quasi-experimental settings providing variation in students' school assignment which is independent of school quality. I first estimate indicators of school quality using the students not changing schools, and in a second step I relate the outcomes of changers to the estimated indicators of school quality of the schools. In the first quasi-experiment I study students changing schools, in the second I study students moving between municipalities and in the last I study changes in neighborhoods' assigned school. In each quasi-experiment there is no evidence that students sort to new school by observables, conditional of the old school. Furthermore, in each quasi-experiment outcomes are related to the old and new school as we would expect if VA represented causal effects of the school attended. Outcomes includes exam scores, and labor market outcomes to the extent that data is available.

The quasi-experiments also allow us to study the relationship between transitory VA and student outcomes. Unlike persistent VA, transitory forecast outcomes of an entrant with bias. However, there are generally significant positive coefficients, indicating that transitory VA also in part capture instruction quality. Teacher grades represent a special case, as students don't benefit from transitory VA.

This paper makes several contributions to the VAM literature. First, it demonstrates how it can be possible to construct informative VA indicators even without data on lagged achievement or the school assignment mechanism. While almost all VA literature controls for lagged achievement, Angrist et al. [2020] stress that the estimators they propose can be calculated even with outdated and missing data on lagged achievement. However, their estimators require some oversubscribed schools, and data on the assignment process. In contrast, the estimators I study can be constructed using only data with results at graduation, as well as time-invariant data on familiy background. This is useful in a setting like the current paper, where lagged achievement data may not be available, and it will be a long time from the introduction of any testing scheme before it is possible to study VA using lagged achievement. However, it can also be useful to study impacts of early school quality, even if pre-school achievement is not recorded.

Second, I study VA throughout compulsory school. Previous studies typically study VA during a year or some stage between tests. However, VA estimates that do not require lagged achievement data allow us to study the entirety of compulsory school and to study the effect of school quality at different stages, similar to what Carneiro et al. [2021] do for the timing of parental earnings. Interestingly, and in contrast to Heckman and Carneiro [2003], I find that late school quality matter most for later outcomes.

Third, I distinguish between persistent and transitory VA. Persistent VA, estimated as by Chetty et al. [2014b], evolves gradually and predicts exam score without bias. Transitory VA is unexplained performance net of persistent VA. This can be instruction quality, which will impact an entrant to a school, or unobserved characteristics of the students, which will not. The significant but smaller than one-to-one relationship between transitory VA and entrants' outcomes strongly suggest that transitory VA reflects both of these. Furthermore, from the dispersion of estimated transitory and the relationship between transitory VA and the outcomes of an entrant, we can conclude that, at least in Norwegian compulsory schools, instruction quality both have a substantial persistent school-level component, and a more volatile component, which may reflect individual teacher quality.

Fourth, recent years have seen increasing interest in VA indicators based on non-test outcomes. Jackson [2018], Jackson et al. [2020] find that non-test school quality is even more important for longer-term outcomes than schools' effects on test scores. As teacher grades arguably reflect a broader set of skills, including e.g. classroom participation, Norwegian teacher grades have previously been used as measures of non-cognitive ability (Falch et al. [2014]). However, the much weaker relationships between VA based on teacher grades and other outcomes illustrate the challenges inherent in using teacher grades to evaluate schools. Despite teacher grades being highly predictive at the individual level, differences in grading practices may mask quality differences between schools. This is likely to be the case for any measure that requires the teacher to evaluate student outcomes in a non-schematic way.

Finally, the current study links learning outcomes and long-term outcomes. When studying school quality or when interventions in the schooling system are evaluated, results are usually in the form of an effect on learning outcomes (e.g. Roland G. Fryer [2017], Angrist et al. [2020]). However, the motivation is often, at least in part, a belief that improvements in school will also promote longer-term outcomes. This study connects learning outcomes to long-term outcomes of interest to policy makers, similar to Chetty et al. [2011] do using Project STAR. It does so using general variation in school quality, suggesting that the (implied) effect of learning on long-term outcomes may be generally relevant (as opposed to e.g. very specific interventions, that may impact strongly on either learning or long-term outcomes, depending on their exact design).

The remainder of the paper proceeds as follows. In Section 2 I describe the institutional context and and data. In Section 3 I present the empirical approach. In Section 4 I present the estimated VA indicators and associations with short- and long-term outcomes, and in Section 5 I present the results from the quasi-experiments. The final section concludes.

2 Institutional setting and data

2.1 Compulsory education in Norway

Compulsory education in Norway lasts for 10 years and is divided into primary (grades 1-7) and lower secondary (8-10) schools. The school system is almost exclusively public, with less than

5 percent of compulsory school students attending private schools.[1] Students are assigned to a school by the municipality based on residence, and most students attend their neighborhood school. In some cases, parents may have the option of choosing a different school than the neighborhood school, but this will subject to capacity

Norwegian schools don't have grade teachers. Teachers will often teach students in different grades and tend to follow the same students within the major divisions of the school system. In the first years of compulsory schools teachers tend to be generalists, teaching a class in all or most subjects, while later in compulsory school teachers will typically have a limited number of subjects in which they teach students from different classes.[2]

Since 2007, students in grades 5 and 8 take national standardized tests in literacy, numeracy, and English. Since 2010, students in grade 9 have taken the same tests in literacy and numeracy as the grade 8 students. These tests are taken early in the academic year, and are often considered exit scores from the previous grade. At the end of compulsory school students get teacher grades in about 11 subjects, and sit one oral and one written exam. The average of these grades constitutes the student's GPA, which matters for admission to upper secondary school.

When choosing upper secondary school, students choose between five academic tracks (leading to a diploma qualifying the student for higher education) and eight vocational tracks (leading to vocational diplomas). Students are entitled to at least three years of upper secondary school in one of their three preferred tracks. However, students compete for places based on their GPA, and are not guaranteed to get their preferred track or school. Thus, unless a student knows that his preferred track and school will be under-subscribed, teacher and exam grades at the end of compulsory school will be high stakes.

While almost all students enroll upper secondary (about 98 percent enroll directly after finishing compulsory school), drop-out and delayed graduation is considered a serious problem. Nominal duration of upper secondary is 3-4 years, but only about 75 percent graduate within

[1] Most private schools are funded by the government about similarly as public schools. These schools are only allowed to charge limited tuition fees. For-profit schools are not allowed; in order to operate a private school the school must be a religious or pedagogical alternative to the public schools. Less than than 0,5 percent of students attend international schools not funded by the government.

[2] E.g., a teacher in lower secondary may teach the same students in a limited number of subjects from grade 8 to 10, possibly at the same time also teaching other students in other grades in the same subjects, and then start over with a new group of grade 8 students when the older students graduate from grade 10.

five years.

2.2 Data on student background and outcomes

The population in the analyses consists of compulsory school students graduating in the years 2002-2019. Figure A1 in the Appendix show the number or students per cohort, which mostly varies around 60,000 students. The data used in this paper are administrative data on standardized tests and end of compulsory school grades for the entire student population. The former is available from 2007 (2010 for the 9th grade test) and the latter from 2002. In the following I will index students by their (end of compulsory) graduation year. Thus, while exam scores and teachers grades are available from 2002-2019, the 2010 graduation cohort is the first for whom 8th grade test exist, and the 2012 (2013) cohort the first for whom I observe the 9th (5th) grade test. Within the cohorts for whom tests are observable, few students have missing values (5-10 percent for each outcome, except for the 5th grade test, which is missing for 10-15 percent), as shown in Figure A2 in the Appendix. To simplify interpretation, exam scores, teacher grades and test scores are standardized to have mean zero and standard deviation of one within each cohort.

Students are linked to parents to construct measures of student background, including the student's gender, immigration background, residential address, and the parents' highest level of education. Figure A3 shows the evolution in the share of female students, students with at least one parent with higher education and the shares of students that are immigrants or Norwegian-born with two immigrant parents. The share of highly-educated parents has increased steadily, from about 40 percent for the 2002 graduates to 54 percent for the 2019 graduates. The share immigrant students increased before decreasing again, and is 7 percent in for the 2019 students, while the share of Norwegian-born children of immigrants has increased from 1.8 percent to 6.4 percent.

Students are also linked to long-term outcomes, including data on progression through and eventual completion of upper secondary, completed years of schooling and labor earnings. Post-compulsory school outcomes are measured up to or in 2017 (except completion of high school, which is also observed in 2018), i.e. 15 years after the first cohort graduates from compulsory school, and when these students are about 31 years old. As completed education and earning are taken from population-wide administrative data, outcomes are observed for almost all students (as is shown in Figure A4). The only outcome strictly limited by data availability is high school completion,[3] which is measured five years after graduation from lower secondary, and thus is observable for cohorts graduating in 2013 or earlier.

However, as Norwegian students often complete their education relatively late, this limits the scope for analyses of completed schooling and in particular earnings. Figure A5 show how average long-term outcomes vary between cohorts. About 51 percent of 2011 graduates and 24 percent of 2007 graduates are registered as in education in 2017, and there are substantial average earnings differences between cohorts. NEET status (not in employment, education or training) is however fairly constant at 13-15 percent for 2012 and earlier graduates. Figure A6 shows the relationship between earnings and standardized compulsory school GPA by graduation cohort. While the relationship is positive for all cohorts, it is much flatter for more recent cohorts, and close to zero for the 2007 graduates. Earnings from before about age 30 are not strongly predictive of life-time earnings (Bhuller et al. [2017], Kirkeboen et al. [2016]).

The first students to sit the 8th grade test graduate from compulsory school in 2010. Thus, while we can follow several cohorts of students with grade 8 tests into upper secondary, we can only study completion of upper secondary for the first three cohorts, and the 2013 cohort is the only cohort for whom we observe 5th and 8th grade test, end-of-compulsory exams and grades as well as high school completion. The cohorts with data on 8th grade or earlier test scores and informative measures of long-term earnings outcomes do not overlap at all.

3 Empirical approach

In this section I lay out a simple model for measurement of school quality which relates estimates of secondary school quality that control for primary school results to those that do not.

School results are observed at the end of two periods 0 and 1, corresponding to primary and lower secondary school. Results in primary school z_0 depend on quality in primary school Q_0, student characteristics $x\beta^0$, and an idiosyncratic error term:

[3] Academic tracks last three years. Vocational mostly last four years, but some courses last longer. A substantial share of students change track, in particular from vocational to academic.

$$z_0 = Q_0 + x\beta^0 + \epsilon_0$$

Allowing for some persistence in results from primary to lower secondary, captured by the coefficient λ, results in lower secondary school z_1 can be expressed as a function of previous results, school quality in lower secondary, and student background;

$$\begin{aligned} z_1 &= \lambda z_0 + Q_1 + x\beta^1 + \epsilon_1 \\ &= \lambda Q_0 + Q_1 + x(\lambda\beta^0 + \beta^1) + (\lambda\epsilon_0 + \epsilon_1), \end{aligned}$$

where the second equality makes clear that we can substitute for previous results z_0 to express z_1 as a function of school quality in primary and lower secondary and student background characteristics.

I assume that the error terms ϵ_0 and ϵ_1 are independent with expectation zero, and also uncorrelated with school quality and observed characteristics. With these assumptions, the difference between observed results in lower secondary and results expected from the students' background and previous results reflect school quality in lower secondary:

$$Q_1 = E[z_1 - \lambda z_0 - x\beta^1] \tag{1}$$

Eq. (1) is the traditional VA measure of school quality used by a range of previous studies and constructed by controlling for previous results. Alternatively, conditioning on student characteristics but not previous results, we get an average school quality across primary and lower secondary, where quality in primary school is weighted by its persistence in determining results:

$$Q_{av} = \lambda Q_0 + Q_1 = E[z_1 - x(\lambda\beta^0 + \beta^1)] \tag{2}$$

3.1 Estimating school quality

I follow Chetty et al. [2014a] and estimate school-by-year value-added, Q_{st}, by adjusting students' results, z_{ist}, for a vector of covariates, x_{ist}:

$$z_{ist} = Q_{st} + x_{ist}\beta + \epsilon_{ist} \tag{3}$$

Here, z_{ist} represent the results (typically exam or test scores) of student i graduating from school s at time t.[4]

The vector of covariates will always include a cubic in socioeconomic index, $X_{ist} = \tilde{x}_{ist}\hat{\beta}$, as well as a school-average value of this index. To construct the index I regress exam score on a set of dummies for gender*immigration status*socioeconomic status (five categories based on parental education) and the combination of the levels of parents' highest completed educations, and get the predicted exam score for each student.[5] Other than the socioeconomic index and and the school-level average socioeconomic index, the set of controls always includes graduation year. Some specifications also include a cubic in the grade 8 test score (average of available tests), as well as the school mean for the average grade 8 test score.

As emphasized by equations (1) and (2), whether I control for previous results or not changes the interpretation of the VA indicators. Controlling for results from primary school gives a VA indicator for lower secondary school quality, as in (1), while controlling for background characteristics only gives a composite measure of quality for both primary and lower secondary, as in (2). While most previous studies have focused on value-added indicators controlling for previous test scores, I will mostly focus on indicators controlling for family background. Thus, the quality experienced by cohort t will be the total quality throughout compulsory school.

From estimating equation (3), I obtain estimated school-by-cohort residuals by taking aver-

[4] As a large majority start school the year they turn six and grade retention is almost non-existent, graduation cohorts closely corresponds to birth cohorts.

[5] For the construction of the VA indicators there is no need to summarize socioeconomic background in terms of an index; all observed characteristics could have been included as separate controls in the analyses. However, in the quasi-experimental analyses presented in Section 5, sample sizes are much smaller, making it necessary to reduce the dimensionality of the controls. Summarizing socioeconomic background in an index also facilitates analyses of whether and how students sort to schools.

ages of individual-level residuals:

$$\hat{Q}_{st} = \bar{\hat{\epsilon}}_{\cdot st} = \bar{z}_{\cdot st} - \bar{x}_{\cdot st}\hat{\beta}$$

Still following Chetty et al. [2014b], I estimate persistent value-added by a shrinkage estimator. Expected school quality for a given cohort in a given school is predicted using estimated school-by-cohort residuals from other cohorts, allowing for drift in school quality. I.e, given $\boldsymbol{Q}_{s,-t} = (\hat{Q}_{s1}, ..., \hat{Q}_{s,t-1}, \hat{Q}_{s,t+1}, ..., \hat{Q}_{sT})$, expected school quality for cohort t is predicted as follows;

$$\hat{\mu}_{st} = E[Q_{st}|\boldsymbol{Q}_{s,-t}] = \boldsymbol{Q}_{s,-t}\hat{\rho}$$

where $\hat{\rho}$ is an estimated auto-correlation vector, which may depend flexibly on time difference, and thus captures persistence in school results. In contrast to Chetty et al. [2014a] I find that the correlations are rather stable, almost irrespective of time difference, at .2-.3 (lower when controlling for previous test scores). This is similar to the long-term correlation of Chetty et al. [2014b], but smaller than the short-term correlations. A likely explanation is that Chetty et al. [2014b] study teacher quality, which may be more persistent in the short term. School quality on the other hand, will change as different cohorts are taught by different teachers. However, although school quality varies more from year to year, there is still a stable component to it, reflecting some shared aspects of the school, over and above individual teachers. Because of this stability of the auto-correlation vector, I will only estimate auto-correlations for two lags, and then use the value for the second lag also for greater time differences in the following analyses (similar to the procedure of Chetty et al. [2014b], but with shorter lags adapted to the stable correlations).

I also estimate school-by-cohort residuals net of persistent differences:

$$\hat{\eta}_{st} = \hat{Q}_{st} - \hat{\mu}_{st}$$

While $\hat{\mu}_{st}$ captures the persistent (although possibly gradually drifting) quality of school s as experienced by cohort t, $\hat{Q}_{st}$ captures the unexplained performance of cohort t. Thus, $\hat{\eta}_{st}$ captures average value-added of school s for cohort t over and above the persistent quality, and will reflect contributions of individual teachers (as teachers assigned typically vary across cohorts), characteristics of the student cohort, and student-teacher match.

Based on the definition of school quality and previous research (e.g. Chetty et al. [2014a]), we expect $\hat{\mu}_{st}$ to be reflected in the school results of a random student entering school s and graduating with cohort t. Whether $\hat{\eta}_{st}$ is similarly reflected is an empirical question, depending on whether $\hat{\eta}_{st}$ mostly reflects teacher characteristics (which should impact on the results of the entrant) or characteristics of the other students (which, absent peer effects, will not affect a randomly placed student).

I will construct several measures of school quality, using exam scores or teacher grades as outcome measures. I will mostly focus on exam scores, as these avoid problems with differences in grading practices associated with teacher grades. For recent cohorts it is possible to condition on grade 8 test scores. However, for earlier cohorts this is not possible, and control variables will be restricted to the socioeconomic index.

3.2 Evaluating effects of school quality

Having estimated persistent and transitory school VA ($\hat{\mu}_{st}$ and $\hat{\eta}_{st}$) I will next study associations between estimated persistent VA and short-term (exams, teachers grades) and long-term outcomes (further education, earnings). I regress each outcome y_{ist} of a student i graduating from school s at time t on estimated school quality, controlling for student and school characteristics x_{ist}:

$$y_{ist} = \gamma\hat{\mu}_{st} + \theta x_{ist} + \nu_{ist} \tag{4}$$

The controls include the socioeconomic index (X_{ist}), school*cohort means of the index, and year dummies.

$\hat{Q}_{st}$, and also $\hat{\eta}_{st}$, will depend on the residuals ϵ_{ist} of students graduating from school s at time t, and must be expected to be correlated with residuals ν_{ist} in other outcome equations for these students. $\hat{\mu}_{st}$ on the other hand, is predicted from from $\boldsymbol{Q}_{s,-t}$, which is related to ϵ_{ist} only through persistent school differences. Interpreting unexplained persistent result differences between schools as reflecting school quality thus requires the assumption that $cov(\hat{\mu}_{st}, \nu_{ist}) = 0$. The coefficient γ measures the ability of the estimated VA indicators to forecast average outcomes. I will follow Chetty et al. [2014b] and denote the VA indicators as (forecast) unbiased if γ=1, i.e. if the indicators on average perfectly forecasts outcomes.

However, there can also be persistent differences between schools not reflecting school quality. The analysis above does not distinguish between a school and the students at this school. Thus, if there are differences between schools in students' unobserved characteristics, these differences will be interpreted as school quality. Unobserved differences in student composition may arise e.g. because of residential sorting combined with neighborhood schools, and thus be unrelated to school quality. This can give rise to $cov(\hat{\mu}_{st}, \nu_{ist}) \neq 0$.

To rule out such a correlation I will draw on variation from three quasi-experiments: School changers (students observed at two different schools), movers (students moving between municipalities), and school district changes (neighborhoods changing local schools). In each of these quasi-experiments the original association between neighborhood and school assignment is broken. Thus, the student is further distanced from the outcomes of the students in other cohorts used to estimate school quality. This potentially reduces correlations between unobserved persistent characteristics and measured school quality and thus allows estimating the effect of school quality on long-term outcomes. I will discuss the validity of the quasi-experiments further when presenting the results.

Given that the quasi-experiments are valid, they also make it possible to study the effect of school-by-cohort value-added, $\hat{\eta}_{st}$. $\hat{\eta}_{st}$ will depend on the residuals ϵ_{ist} of the students used for estimating school quality. However, with valid quasi-experiments it is possible to estimate persistent school quality and school-by-cohort value-added from the stayers (students or neighborhoods) not changing school, which will be independent of ν_{ist} for the students that do change. Thus, we can estimate the separate effects of persistent and transitory school value-added, estimated from the stayers, on the outcomes of the movers (students or neighborhoods that do change school attended or assigned to).

4 Associations between value-added and students' outcomes

In this section I start out by estimating and briefly presenting the estimated VA indicators. I next investigate whether the VA indicators are able to forecast exam scores, and whether VA

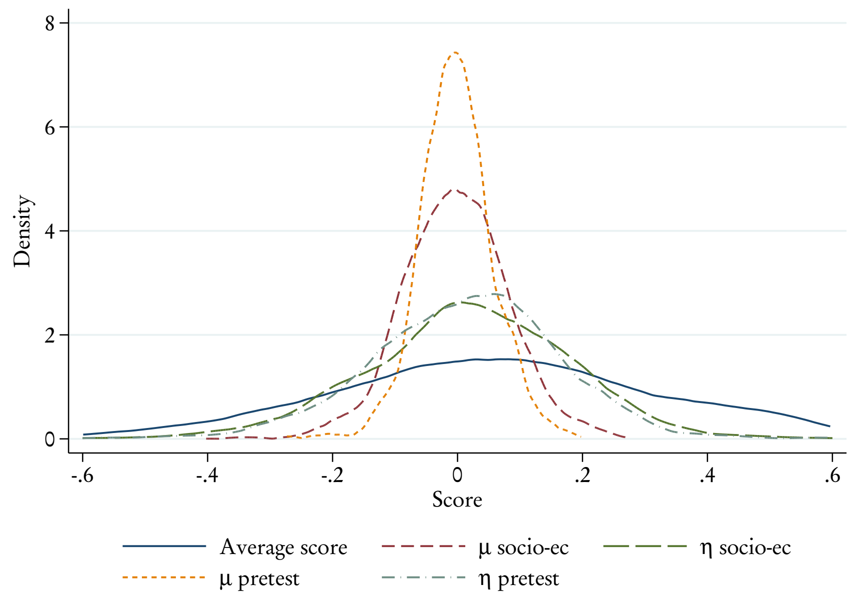


Figure 1: School average exam score and estimated measures of school quality for the 2014 cohort.

indicators are also predictive of longer-term outcomes.

4.1 School quality

I start out by estimating school quality for the 2010--2014 cohorts. The 2015--2019 graduation cohorts will be reserved for testing the indicators. The quality measure is end-of-compulsory school exam scores.

Figure 1 shows the student-weighted distributions of school mean exam scores and estimated school quality measures for the 2014 students. School quality estimates always control for socioeconomic background, and additionally the grade 8 standardized test when indicated. For both measures of persistent VA (μ) I also present the unexplained, transitory performance (η).

We see that there is much less variation in either of the school quality measures than in average exam scores. This reflects that, even in a relatively equal society like Norway, with limited residential sorting and almost no private schools, a large part of differences between schools reflect student composition. While the student-weighted standard deviation of average exam score is 0.29 student-level standard deviations (SDs), the standard deviation of persistent VA adjusted for pretest is 0.085 SDs. There is less variation in estimated persistent school quality when controlling for pretest (0.060 SDs). Finally, there is greater variation in transitory performance than in persistent quality, and transitory performance varies more when not controlling for pretest (0.19 vs 0.17 SDs). While clearly not identical, the indicators controlling for different sets of covariates are highly correlated, with correlation coefficients of 0.83 for persistent VA and 0.77 for transitory performance.

It is possible to construct similar indicators for any outcome variable. In Figures A7 and A8 in the Appendix I show similar distributions of estimated school quality measures based on standardized tests in grades 8 (Figure A7) and 5 (Figure A8). While the estimation is entirely analogous to that of the indicators in Figure 1, there are some notable differences in interpretation. First, the earliest measure of performance in school is in grade 5. Thus, any measure of school quality up to grade 5 can only adjust for family background. Second, as standardized tests in grades 5 and 8 were introduced in 2007, the cohort completing lower secondary in 2013 is the first to sit the grade 5 test. Thus, the indicators based on grade 5 scores and the indicators based on grade 8 scores that control for pretest are only estimated from the two cohorts completing lower secondary in 2013 and 2014. Finally, the indicators in Figures A7 and A8 pertain to lower secondary schools. These schools will in general not correspond to the primary schools that the students actually attended. Thus, these indicators cannot be interpreted as measures of the quality of the schools for which they are estimated. Rather, these indicators measure primary school quality as experienced by the students in a given lower secondary school. While indicators would need to pertain to the school responsible to be relevant for informing stakeholders, this way of constructing the indicators allow us to directly compare the quality distributions of primary and lower secondary education in Norway. As such, we see that these are very similar.

A key question is whether the indicators really capture persistent differences between schools. As noted, I find auto-correlation in Q ranging from .2-.3. In Table 1 I further investigate the out-of-sample predictive power of the indicators. I regress student-level exam scores of students completing compulsory school in 2015 on indicators of school quality constructed from the 2010-2014 cohorts, controlling for the same control variables used in the construction of the indicators but for the 2015 students themselves.

Table 1: Regressing exam performance of 2015 graduates on 2010-2014 school quality

	(1)	(2)	(3)	(4)	(5)	(6)
	Indicators control for previous score			Indicators control for background only		
Exam value-added	1.051**	1.028**		1.065**		
	(0.139)	(0.128)		(0.088)		
8th grade value-added	0.751**		0.661**		0.634**	
	(0.102)		(0.104)		(0.062)	
5th grade value-added	0.382**					0.354**
	(0.050)					(0.054)
Controls (cubic + school mean):						
Grade 8 score		Yes				
Grade 5 score			Yes			
Family background	Yes	Yes	Yes	Yes	Yes	Yes
N students	49253	49253	49253	49253	49253	49253
N clusters	1044	1044	1044	1044	1044	1044
R^2	0.237	0.481	0.414	0.237	0.234	0.231

Note: Outcome is exam score of 2015 cohort. Value-added indicators are constructed from the 2010-2014 cohorts. All regressions control for cubic in an index of socioeconomic background for the sample index as well as the school*cohort mean index (same index as used for constructing indicators) and year dummies. Specifications controlling for cubic in pretest also control for school-mean pretest. Standard error are clustered at the school. Significant at * 10%, ** 5%

As highlighted by equations (1) and (2), the sets of controls used when estimating school quality decides the interpretation of the estimated indicators. Furthermore, regressing exam scores on school quality, controls for students' background need to be consistent with the indicators. In Table 1 I break down the total contributions from school quality and students' background in different ways across the columns.

In column (1) I present the results from regressing exam scores on the most comprehensive specification, controlling separately for estimated school quality before grade 5, from grade 5 to 8 and from 8 to the end of compulsory school, as well as students' (pre-school) background.[6] On average, differences in exam score value-added, estimated controlling for 8th grade scores, predict exam scores without bias. The estimated coefficient on the value-added indicator is precisely estimated and not significantly different from one.

[6]Students' background is partly decided at birth (sex and immigrant background) and partly measured at age 16 (parents' education). However, even if parents' formal education may change during the students' childhood, for most students it does not. Furthermore, in the relationship with their children's school performance, parents' education largely is a proxy for characteristics of the parents that are likely fixed. While the background variables themselves reflect pre-school characteristics, the relationships with exam scores may change over time. As highlighted by eqs. (1) and (2) the coefficients on student background will represent the total association between background and school performance.

Value-added in 8th and 5th grade (8th grade value-added estimated controlling for 5th grade score and 5th grade value-added only controlling for family background) also predict exam scores, conditional on exam value-added and student background. As the 8th and 5th grade value-added are in units of test scores while the outcome is exam score, we strictly cannot speak about "unbiased", as for exam value-added. However, as each of these outcomes is measured in student-level standard deviations the scales are directly comparable. From Figures 1, A7 and A8 we also know that the distributions of the indicators are similar. Thus, we can conclude that 8th grade and even more so 5th grade value-added are less strongly associated with exam scores than exam score value-added, with a one SD difference in value-added being associated with exam score differences of 0.75 SD and 0.38 SD, respectively.

In column (2) I disregard school quality in primary school, and rather control directly for the end result: The students' own 8th grade score. This substantially increases the explanatory power of the regression. However, it leaves the coefficient on exam score value-added essentially unchanged, and not significantly different neither from the coefficient in column (1) nor from one. Similarly in column (3), the coefficient on 8th grade value-added is not significantly different from the coefficient in column (1) when I control for 5th grade score and disregard school quality in lower secondary and before 5th grade.

Columns (4)-(6) similarly regress exam scores on indicators constructed only controlling for socioeconomic background. Consistent with how the indicators are constructed, these specifications only control for students' background, and not previous test scores. All the associations between the indicators and exam scores are very similar to those in column (1), and again the exam value-added indicator predicts exam scores without bias.

The results in Table 1 clearly demonstrate that the indicators have predictive power out-of-sample. They thus capture some persistent element of school results over and above student composition. However, 2015 is close to the estimation period. In Figure 2 I present results from regressing exam scores on value-added, similar to columns (2) and (4) of Table 1, for each of the years 2010--2019. Unsurprisingly, for each year of the estimation period the coefficients are close to one. After the estimation period the coefficients gradually decline, and while they stay close to one in the first two years, by 2018 the coefficients are significantly different from one. However, at this time, four years after the end of the estimation period, the coefficients are still

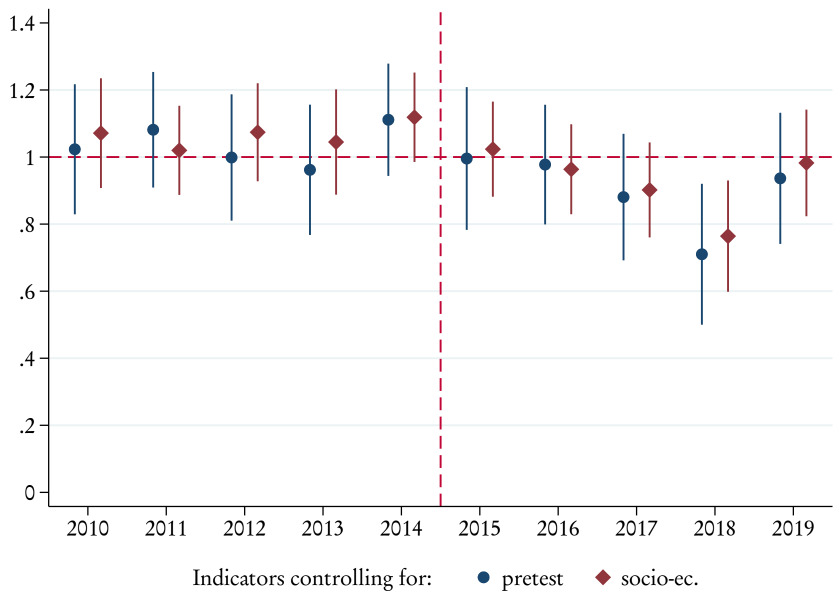


Figure 2: Exam score and estimated school quality by cohort
Note: The graph show the estimated relationship between exam score and VA indicators, similar to columns (2) and (4) of Table 1. VA indicators are based on the 2010-2014 cohorts.

0.7-0.8, and the confidence intervals overlap with the coefficients of the previous years. Also, in 2019, the final year of the sample, the coefficients revert to one. Detailed results for 2019 similar to those in Table 1 are presented in Table A1 in the Appendix.

4.2 Long-term student outcomes

The next question is whether the value-added indicators capture variation in competence that are restricted to exam scores, or to what extent school value-added also predict students' later outcomes. In order to study long-term outcomes, the students must have reached sufficient age such that the outcomes are realized. This severely restricts both the samples and the types of indicators that are relevant. As discussed in Section 2, completion of upper secondary, completion of higher education, and labor market participation and earnings cannot be measured reliably until ages 21--30. For higher education, labor market participation, and earnings, I will study cohorts graduating from compulsory school in 2002 and 2003, who will be 30-31 years old in 2017. As even the earliest available cohorts are only just relevant for studying earnings, I will relate the outcomes of the 2002 and 2003 cohorts to persistent value-added estimated from the 2004-2008 cohorts. For these cohorts, standardized test scores are not available, and I will therefore only be able to study VA indicators controlling for family background.

Table 2 shows associations between estimated VA and medium- and long-term outcomes. Each cell reports the key coefficient from a separate regression, regressing an outcome variable on a VA indicator or student in-school outcome, controlling for the index of family background used in constructing the value-added indicator. Each column represents a different outcome variable, while each row represents a measure of VA or student performance. The variable of primary interest is the value-added indicator constructed from exam scores, in the first row. However, I also report associations with VA estimated from oral exams and teacher grades, and, to help interpretation of the magnitudes of the associations, the individual-level cross-sectional associations between the different outcomes and students' exam scores and teacher grades.

The first cell of the first row of Table 2 reports the ability of the indicator based on exam score to predict school mean exam scores out of sample. As we would expect from the results of the corresponding analysis for later cohorts shown in Table 1, I find a coefficient close to one, although slightly attenuated. I.e., the exam-based VA indicators forecast average exam scores out of sample with little bias. In the second and third columns we see that there is not a one-to-one relationship between the written exam-based VA indicator and oral exam scores and teacher grades. However, the association between the VA indicator and oral exam scores or teacher grades is similar to the the corresponding individual-level relationships (shown in the fifth and sixth row).

There is consistently a highly significant and strong association between the written exam-based indicator and later outcomes. The next three measures, on-time completion of the first year of upper secondary, graduating from upper secondary within five years, and completed years of schooling, are all more strongly associated with the exam-based indicator than with own exam score, although the difference is not significant for years of schooling.

In the next columns I show results from similar analyses of longer-term outcomes related to earnings and labor marked participation. A potential challenge studying these outcomes is that more academically successful students stay longer in school, which may influence measurement of the outcomes. In column (7) we see that this is indeed the case. 11 percent of all students

Table 2: School quality and short- and long-term outcomes

	(1) Written exam	(2) Oral exam	(3) Teacher grade	(4) Completed year 11	(5) Completed high school	(6) Years schooling	(7) In education	(8) Employed	(9) NEET	(10) Earnings	(11) Log earnings
$\hat{\mu}$ written exam	0.850**	0.508**	0.564**	0.162**	0.247**	0.970**	-0.007	0.081**	-0.061**	1.158**	0.160**
	(0.058)	(0.053)	(0.059)	(0.025)	(0.024)	(0.163)	(0.013)	(0.016)	(0.014)	(0.128)	(0.022)
$\hat{\mu}$ oral exam	0.496**	0.809**	0.791**	0.030	0.035	0.560**	0.021	0.053**	-0.046**	0.652**	0.084**
	(0.064)	(0.056)	(0.062)	(0.024)	(0.024)	(0.181)	(0.014)	(0.016)	(0.014)	(0.132)	(0.023)
$\hat{\mu}$ teacher	0.269**	0.413**	0.915**	0.017	0.069**	0.285**	0.005	0.020*	-0.022**	0.457**	0.071**
	(0.044)	(0.039)	(0.039)	(0.018)	(0.017)	(0.122)	(0.009)	(0.011)	(0.010)	(0.088)	(0.016)
Written exam score	1.000**	0.500**	0.617**	0.114**	0.143**	0.902**	0.015**	0.032**	-0.031**	0.521**	0.078**
	(0.000)	(0.004)	(0.004)	(0.002)	(0.002)	(0.015)	(0.001)	(0.001)	(0.001)	(0.013)	(0.002)
Oral exam score	0.484**	1.000**	0.587**	0.107**	0.141**	0.848**	0.013**	0.035**	-0.033**	0.556**	0.084**
	(0.004)	(0.000)	(0.003)	(0.002)	(0.002)	(0.015)	(0.001)	(0.001)	(0.001)	(0.013)	(0.002)
Teacher grade	0.709**	0.696**	1.000**	0.178**	0.230**	1.300**	0.015**	0.059**	-0.054**	0.771**	0.113**
	(0.005)	(0.003)	(0.000)	(0.002)	(0.002)	(0.015)	(0.001)	(0.002)	(0.002)	(0.012)	(0.002)
N	83371	83371	83144	82871	82871	81317	83371	83371	83371	81297	71667
# clusters	2025	2025	2024	2025	2025	2024	2025	2025	2025	2024	2005
$\bar{y}$	0.015	0.015	0.037	0.808	0.714	13.784	0.110	0.860	0.114	4.407	1.512

Note: Each cell is a separate regression of outcome on VA indicator or exam/teacher grade on the 2002 and 2003 compulsory school graduation cohorts. Outcomes (1)-(3) are from the end of compulsory school, (4) is observed one year after completing compulsory school and (5) five years after. Outcomes (6)-(12) are observed in 2017, i.e. 14-15 years after graduation from compulsory school, around age 30. (6) is nominal duration of highest completed degree (in years, including compulsory school); (7) is a dummy for whether the person in education in 2017; (8) is an earnings-based employment measure (earnings $>$ G, approx USD 10 000); (9) is a dummy for not in employment, education or training; (10) is annual labor earnings and (11) is log annual earnings. The indicators are constructed from the 2004-2008 cohorts. All regressions control for cubic index of socioeconomic background (same as used in indicators), school*year mean index and year dummies. Standard errors are clustered at the school level. Significant at * 10%, ** 5%

are still under education 14-15 years after completion of compulsory schooling, and students with higher exam scores more often. However, the VA indicator is unrelated to whether the student still is in education. The VA indicator is however related to labor market outcomes. The share employed[7] is higher among students from high-VA schools and inactivity (NEET; not in employment, education or training) is less common. Finally, both for average earnings and for log earnings (for the sub-sample with earnings above the cut-off) the association with exam-based indicators is about three times stronger than the associations with individual-level exam scores.

A school-level one-standard deviation difference in exam value-added (i.e. a .085 student-level SD difference), corresponds to a predicted difference of 2.8 percentage points in upper secondary completion, a 0.9 percentage points difference in labor market participation and a 1.5 percent earnings difference (given positive earnings). These associations are strong relative to the individual-level cross-sectional associations, and the differences in secondary school completion and participation are also substantial relative to the baseline levels reported in the last row of the table. This suggests that schools may play an important role in providing skills that have a lasting impact, and that exam performance measures this contribution in a relevant way.

Table 2 also reports associations between outcomes and indicators constructed from oral exam scores and teacher grades. The association between the oral-exam indicator and oral exam score and the teacher-grade indicator and teacher grades are both strong, similar to exam-score VA and exam scores. However, with the exception of a strong association between oral-exam VA and teacher grades, the associations with other in-school outcomes are much weaker. Also, indicators based on teacher grades are not as strongly related to average exam scores as indicators based on exam scores are to teacher grades. The associations between post-school outcomes and teacher grade-VA indicators are consistently weaker than the associations between the same outcomes and exam-indicators. The oral exam-indicators are mostly in-between.

Thus, indicators based on written exam scores are predictive both out of sample and in other domains. The indicators capture persistent differences in terms of school performance that are not explained by students' characteristics. Furthermore, differences between schools

[7] Employment is measured as earnings greater than the basic amount of the Norwegian social security system, about USD 10,000. This is often used as a measure of labor market participation. An alternative measure based on the reported percentage of a full-time position gives similar results.

in exam performance are also reflected in the students' later outcomes, including labor market participation and earnings.

Indicators constructed from oral exam scores or teacher grades also capture persistent differences between schools. These indicators are also predictive of later outcomes, but much less so than indicators based on exam scores, in particular teacher grade VA. This is despite the fact that student-level teacher grades predict later outcomes better than exam scores, suggesting that while teacher grades are informative at the individual level, there are school-level biases if we want to study differences in school quality, e.g. differences in grading practices.

A potential mechanism behind the associations could be that students from high-VA schools get better grades and thus get admitted to better upper secondary schools. However, students compete for places in upper secondary based on their grade point average, which is mostly based on teacher grades. Thus, the much weaker association between teacher-grade VA and later outcomes suggests that this mechanism is not very important.

High- and low-VA schools may be located in different labor market, such that differences in later outcomes do not reflect VA. In Table A2 in the Appendix I reproduce Table 2, but with municipality fixed effects. The results are slightly attenuated, but qualitatively similar to those reported in Table 2. By including municipality fixed effects I disregard between-municipality variation in estimated VA. However, as municipalities are responsible for compulsory schools, this may remove relevant variation. Furthermore, many Norwegian municipalities are small. 178 municipalities, with 14 percent of the students, only have one single school, and thus do not contribute to the fixed effects estimates. Thus, the fixed effect analysis may underestimate the association between school quality and later outcomes.

For most longer-term educational and labor market outcomes I am constrained to use cohorts before the introduction of 5th and 8th grade test score. However, the for the 2013 graduates I have both data on test scores and data on completion of high school. In Table A4 I show how high school completion is related to VA for different stages of compulsory school, similar to what I did for exam scores in Table 1.[8] VA indicators for exam score, 8th grade test score and 5th grade test score all predict completion, for each permutation of VA indicators and controls. VA

[8]In Table A4 VA is estimated from the 2015-2017 cohorts. Table A3 I demonstrate that the associations between VA and exam scores correspond to those for the 2015 and 2019 cohorts in Tables 1 and A1.

later in compulsory school is more strongly related to completion than VA from earlier stages, like for exam scores in Table 1, However, the gradient is less pronounced than in Table 1, and insignificant and close to zero for VA that condition on lagged achievement.

5 Quasi-experimental evaluation of the effect of school value-added

From the previous section I conclude that there are persistent differences between schools in terms of exam performance, which are not explained by observed student characteristics. Furthermore, these differences also predict later outcomes, including labor market participation and earnings. In this section I discuss whether the differences between schools' average outcomes are actually reflecting school quality, or other unobserved differences.

An important concern when trying to disentangle school quality from student characteristics is related to the potential for systematic sorting of students to different schools. If such sorting is present, students attending the same school may share unobserved characteristics that may confound the analysis. In Norwegian compulsory school, the vast majority of students attend their local neighborhood school. Thus, any sorting of students and bias from confounding variables is likely to operate through the students' neighborhood. To address this concern, we would ideally have an experiment where students are randomly assigned to schools, independently of in which neighborhood their families choose to live and thus of characteristics correlated with this choice. As there is no assignment to schools by lottery in Norwegian compulsory education, I will rather rely on three different quasi-experiments, where students in different ways change their actual or predicted school. The first two use students changing schools and/or moving, while in the last assignment of neighborhoods to schools change.

5.1 School changers

Since 2007, students' school assignment is observed when the students sit the standardized tests in grades 5 and 8, and since 2010 also in grade 9. This enables us to observe the students' school assignment at several times throughout compulsory school. In particular, the tests in grade 8 and 9 allow identification of students changing school early in lower secondary. Some students are observed at one school for the grade 8 test and a different school in grade 9; I will refer to these as school-changers. The remaining students, the non-changers, are observed in the same school in both years. In a first attempt to address the potential correlation between students' residuals and schools value-added, I construct value-added measures based on the non-changers who also graduate from the same school as they sit the 8th grade test and study how the value-added of the changers' new school predicts the changers' outcomes, conditional on the changers' old schools. If the value-added of the new school is unrelated to the residual of the student (i.e., $cov(\hat{\mu}_{st}, \nu_{ist}) = 0$ in eq. (4)), conditional on the old school, this will give a consistent estimate of the effect of the value-added of the new school.

In Table 3 I present the results from regressing outcomes of changers on persistent ($\hat{\mu}$) and transitory ($\hat{\eta}$) VA of the old (grade 8) and new (grade 9) school, constructed from non-changers. All regressions also include controls for average characteristics of the students' cohorts in the old and new schools.

The upper panel presents results using indicators controlling for lagged achievement, while the lower panel use indicators controlling only for family background. In column (1) I present results from regressing exam scores on student and school characteristics. In the upper panel, the coefficient on persistent VA of the new school is 0.88, and highly significant. A coefficient smaller than one is to be expected, as the students change school some time between early 8th grade and early 9th grade, and thus do not spend all of lower secondary in their new school. Furthermore, of those changing school at least once, about 20 percent change again before graduating from compulsory school. Still, while the coefficient is smaller than one, it is not significantly so. For comparison, the coefficient on the persistent VA of the old school is 0.28.

Because school-changers and the non-changers used to estimate transitory VA are separate but concurrent groups, I can also study how transitory VA is related to outcomes of students not used to construct the indicators. We see that transitory VA of the new school is significantly related to exam scores, however, the coefficient of 0.32 means that there is far from a one-to-one relationship. Transitory VA likely capture a wide range of causes of result differences, e.g. quality of individual teachers and unobserved characteristics of the students. A coefficient of 0.32 implies that transitory VA mostly reflects characteristics that don't impact an incoming student, but also that a substantial part of the transitory VA is potentially causal. Furthermore, while transitory VA is less predictive of outcomes than persistent VA, the SD of transitory VA is about twice as large as the SD of persistent VA, such that the relative contribution of transitory VA to outcome differences is greater than the ratio of the coefficients in Table 3. As was the case with persistent VA, transitory VA of the old school is also significantly related to exam scores, but less strongly than the transitory VA of the new school.

In column (2) I present similar results for teacher grades. Exam score VA is significantly related also to teacher grades, although less strongly than to exam scores, as found in Table 2. A notable difference from the results for written exam in column (1) is that there is no relationship between transitory VA of the new school and teacher grades. This likely reflects relative grading in the new school. Columns (3)-(5) present results for longer-term outcomes. There are no significant effects of VA on these. This is largely an issue of precision, and the coefficients are not significantly different from the corresponding coefficients in Table 2.

If we are to interpret these findings as causal effects the identifying assumption is that relevant unobserved characteristics of students moving from schools with a given VA are not systematically related to the VA of the new school, conditional on observable controls (including average characteristics of students at the new school). This assumption is not testable. However, we can evaluate its credibility by looking for indications of sorting by observables. In the last column of Table 3 I study how the 8th grade test score is related to the VA measures. While there are strong (and possibly causal) positive relationships between persistent and transitory VA of the new school and exam scores, there are insignificant negative relationships between VA and the pre-determined 8th grade test scores. A lack of a significant relationship is to be expected, given our knowledge of the context.

Historically, data on school quality has not been easily available in Norway. Average end of compulsory school grades have been available since 2002, but data on VA has not been available, and data on transitory VA is not even forecastable.[9] Thus, as there is no indication of systematic sorting of students, I conclude that the results in Table 3 provide credible estimates of the effects of receiving school quality, as measured by VA, on the outcomes of students changing schools.

[9] Compulsory school quality indicators, adjusting end of compulsory school grades for students' family background, were estimated nationally in 2005, but only published for schools in Oslo (Fiva and Kirkebøen [2011]). Since then, school quality indicators have not been available until recently; first for Oslo in Ekren [2015], and nationally in Steffensen et al. [2017].

Table 3: Exam score VA and outcomes of school-changers

	(1) Written exam	(2) Teacher grade	(3) Completed year 11	(4) Completed high school	(5) NEET	(6) Control (pretest/ index)
Indicators controlling for pretest						
$\hat{\mu}^{Old}$	0.284**	0.271**	-0.034	0.165	-0.053	0.141
	(0.116)	(0.123)	(0.083)	(0.134)	(0.072)	(0.142)
$\hat{\mu}^{New}$	0.880**	0.506**	0.053	0.148	-0.108	-0.227
	(0.121)	(0.132)	(0.082)	(0.148)	(0.068)	(0.138)
$\hat{\eta}^{Old}$	0.118**	0.066	0.013	0.083	-0.035	0.023
	(0.046)	(0.046)	(0.029)	(0.055)	(0.026)	(0.050)
$\hat{\eta}^{New}$	0.323**	-0.006	-0.033	0.042	0.031	-0.015
	(0.050)	(0.044)	(0.030)	(0.051)	(0.026)	(0.048)
Indicators controlling for family background						
$\hat{\mu}^{Old}$	0.269**	0.419**	0.065	0.209**	-0.056	0.038
	(0.089)	(0.103)	(0.053)	(0.094)	(0.045)	(0.059)
$\hat{\mu}^{New}$	0.757**	0.548**	0.082	0.185*	-0.050	0.015
	(0.089)	(0.105)	(0.059)	(0.097)	(0.045)	(0.056)
$\hat{\eta}^{Old}$	0.085*	0.087*	0.018	0.078	-0.026	-0.006
	(0.047)	(0.052)	(0.027)	(0.052)	(0.024)	(0.027)
$\hat{\eta}^{New}$	0.325**	0.077	0.012	0.092	0.004	0.005
	(0.046)	(0.048)	(0.027)	(0.057)	(0.022)	(0.025)
Student controls	Yes	Yes	Yes	Yes	Yes	*
N students	8014	7828	6092	2112	6160	8014
N clusters	935	932	904	692	906	935

Note: Each column*panel is a separate regression of an outcome on persistent ($\hat{\mu}$) and transitory ($\hat{\eta}$) VA of the grade 8 and grade 9 schools. Sample is all students with recorded 8 and 9 grade tests (compulsory schools graduation cohorts 2011-). Indicators are constructed from students in the same cohorts that don't change schools. Outcomes in columns (1)-(5) are the same as in Table 2. Outcome in column (6) is the key individual control variable in the regressions in the same panel; pretest in the top panel and the student background index in the lower. All regressions in columns (1)-(5) control for student background by cubic in background index, in the top panel also for cubic in 8th grade test score. The regression in the top panel of column (6) controls for student background, but not test score, the regression in the lower panel neither. All regressions include school-level controls for the grade 8 and 9 schools (school*year average student background, and also test score in the upper panel). Standard error are clustered at the year 8-school. Significant at * 10%, ** 5%.

In the lower panel of Table 3 I show similar results using indicators controlling only for student background, while I control for student background but not lagged in the regression. The results in the last column relate VA to student background rather than the 8th grade test scores. The results are generally very similar to those in the upper panel. In Table A5 in the Appendix I present results similar to those in Table 3, but where I control flexibly for the grade 8 school with school dummies. The results for the new school VA is very similar to those in Table 3.

In Table A6 in the Appendix I repeat the analyses in Table 3 with indicators constructed from teacher grades. While both persistent and transitory VA are significantly related to teacher grades, only persistent VA controlling for student background (and not for lagged achievement) is related to exam scores, and not very strongly so. This suggest that VA indicators based on teacher grades mostly measures differences in grading practices, rather than differences in school quality.

5.2 Movers

For older cohorts, which completed grade 8 before 2008, we are not able to observe school assignment until completion of compulsory school. Thus, we cannot know if they changed school, and cannot directly study school changers as above. However, as school assignment is tied to place of residence we can infer the likely school from the students' address.

In order to create a link between address and likely school, I use the students' neighborhood.[10] The student cohort graduating from compulsory school in 2017 is spread across 11,000 neighborhoods, with 1-88 students in each (average is 6 students). Then, for each neighborhood and cohort of compulsory school students, I find the modal school for students in the neighborhood. As data on school assignment only is available at the end of compulsory school, this will provide a predicted upper secondary school. While some students attend the same school throughout compulsory school, many schools are only primary or upper secondary schools, and

[10] To make "neighborhood" operational I use the students' "basic statistical unit". Basic statistical units are the smallest geographical units used by Statistics Norway for official statistics. Norway is divided into about 14,000 such units, with population in 2017 ranging from 1-6000 (average population is 379). The unit are described as *"small, stable geographical units* which may form a flexible basis to work with and present regional statistics (...) *geographically coherent* areas\ (...) *homogeneous*, with respect to nature and basis for economic activities, conditions for communications, and structure of buildings" (emphasis mine).

many students change school at the transition from primary to lower secondary. Thus, I cannot predict primary school attended. Rather, I will study value-added associated with lower secondary schools, acknowledging that this may in part stem from the contribution of the primary schools previously attended, cf. discussion in Section 3.

From the data on residence, I define two groups of students, movers and never-movers. I define never-movers as the students living in the same municipality throughout compulsory school, while movers are students that move between municipalities at least once during compulsory school. I use the never-movers to estimate value-added based on exam scores and teacher grades, and construct student-weighted average VA for each neighborhood and year. As test score data is not available for these cohorts, I only control for the socioeconomic index.

Table 4 shows the relationships between between movers' outcomes and the persistent and transitory VA associated with the movers' neighborhood at the start of compulsory school and after their first move. In the first column we see that exam scores are strongly related to the persistent VA of the neighborhood after moving, although the coefficient of 0.6 is significantly different from one. However, unlike for school-changers in the previous subsection, I am not able to observe actual school attended. This will cause some measurement error in VA. 60 percent of movers graduate from the modal school of their neighborhood after their first move.[11] Thus, if VA of the school actually attended by the movers is uncorrelated with the VA of their predicted school, we can expect an attenuation bias of 30-40 percent. Adjusting for this bias, the coefficient on persistent VA of the new school neighborhood is close to one. Like in the previous subsection, exam scores are also significantly related to transitory VA, although not one-to-one (also if we adjust for attenuation bias).

In columns (2)-(7) I show how persistent and transitory VA are related to teacher grades and longer-term outcomes. Both persistent and transitory VA of the new school is significantly related to teacher grades, completion of grade 11 and high school and NEET status and earn-

[11]This partly reflects that not all students in a neighborhood attend the modal school and partly repeated moving. 91 percent of never-movers and 70 percent of movers graduate from the compulsory school they are expected to, based on their neighborhood at age 16 and the modal school among the never-movers. 33 percent of movers move more than once. The amount of attenuation bias will further depend on the time spent in the second school before moving again, and whether schools have larger impacts on outcomes at certain ages. It is possible to construct a measure of average predicted VA, based on neighborhood in each year. However, this requires deciding on how to weigh VA in different years together. It is also possible to control for characteristics of schools after the first two. However, as only a minority of students move more than once this will likely be of minor importance, and as subsequent moves may be endogenous to the quality of the second school inclusion of later schools complicates the interpretation of the coefficients on the second school.

ings around age 30 (only at the 10 percent level for NEET and log earnings). Adjusting for measurement error as above, the coefficients on persistent VA are similar to those in Table 2. As in Table 3, transitory VA of the new neighborhood is much more strongly related to exam scores than to teacher grades. For the longer-term outcomes in columns (3)-(7) the coefficients on transitory VA are about 2/5 of the coefficient on persistent VA. VA of the old neighborhood is generally about as strongly related to the outcomes in columns (2)-(7) as VA of the new neighborhood.

For the coefficients on the new neighborhood VA to be informative about effects of a new school, the residuals of the movers must be uncorrelated to the new VA, conditional on observable controls (including their old neighborhood). As in the previous sub-section, I evaluate this by studying whether VA is related to the socioeconomic background index. The last column shows the results. Both persistent and transitory VA are unrelated to student background; the coefficients are insignificant and close to zero (remember that the background index is predicted exam score, and thus have the same scale as exam score). There are however significant associations between background and both persistent and transitory VA of the old neighborhood. The association with background index is also close to the observed association with exam score VA of the old school, in particular for transitory VA.

In the lower panel I present results from regressions with old-neighborhood fixed effects. The results for persistent and transitory VA for the new neighborhood are essentially unchanged. The precision of the estimated relationships between old-neighborhood VA and outcomes is substantially reduced. However, both persistent and transitory VA of the old neighborhood are still significantly related to student background.

The movers are not involved in estimating VA of neither the old nor the new school. The association between old-school VA and student characteristics illustrates how there still may be sorting of students to schools and cohorts within schools. However, in contrast to the observed associations between old-school VA and student characteristics, there is no indication of any corresponding association between new-school VA and student characteristics, conditional on the old school. This matches our knowledge of the context, in particular the general unavailability of data on school quality. Thus, there is no reason to expect significant biases from sorting on unobservables (conditional on the old school and observable characteristics).

Table 4: Exam score-VA and movers outcomes

	(1) Written exam	(2) Teacher grade	(3) Complete year 11	(4) Complete high school	(5) NEET	(6) Earnings	(7) Log earnings	(8) Background index
$\hat{\mu}^{Old}$	0.263**	0.397**	0.093**	0.152**	-0.040*	-0.007	0.027	0.054**
	(0.037)	(0.040)	(0.017)	(0.023)	(0.021)	(0.228)	(0.047)	(0.023)
$\hat{\mu}^{New}$	0.601**	0.378**	0.096**	0.138**	-0.037*	0.773**	0.078*	-0.013
	(0.034)	(0.037)	(0.017)	(0.021)	(0.020)	(0.215)	(0.046)	(0.017)
$\hat{\eta}^{Old}$	0.050**	0.045**	0.012*	0.017*	-0.018**	0.112	-0.006	0.037**
	(0.015)	(0.015)	(0.007)	(0.009)	(0.008)	(0.085)	(0.018)	(0.007)
$\hat{\eta}^{New}$	0.341**	0.085**	0.035**	0.051**	-0.015*	0.194**	0.044**	0.002
	(0.015)	(0.015)	(0.008)	(0.009)	(0.009)	(0.084)	(0.018)	(0.007)
With neighborhood fixed effects:								
$\hat{\mu}^{Old}$	0.259**	0.288**	0.083*	0.125**	0.050	-2.131*	-0.454*	0.120**
	(0.090)	(0.092)	(0.045)	(0.063)	(0.075)	(1.240)	(0.249)	(0.040)
$\hat{\mu}^{New}$	0.592**	0.396**	0.098**	0.144**	-0.070**	0.436	0.049	0.006
	(0.038)	(0.040)	(0.018)	(0.024)	(0.023)	(0.292)	(0.062)	(0.018)
$\hat{\eta}^{Old}$	0.041**	0.028*	0.003	0.004	-0.009	-0.076	-0.087**	0.026**
	(0.017)	(0.017)	(0.008)	(0.011)	(0.011)	(0.162)	(0.034)	(0.008)
$\hat{\eta}^{New}$	0.342**	0.065**	0.034**	0.050**	-0.013	0.213**	0.055**	0.003
	(0.016)	(0.016)	(0.008)	(0.010)	(0.009)	(0.106)	(0.023)	(0.007)
N students	95922	98294	94398	71267	52839	18143	14594	104805
N clusters	10441	10478	10389	9794	9006	6182	5582	10589

Note: Sample is students moving during compulsory school. Outcomes are the same as in Table 3. $\hat{\mu}^{Old}$ and $\hat{\eta}^{Old}$ are persistent and transitory VA (exam scores adjusted for student background) of the modal school lower secondary school of the student's neighborhood when starting school and $\hat{\mu}^{New}$ and $\hat{\eta}^{New}$ are similar VA of the lower secondary school of the student's neighborhood after moving. All columns control for cohort and neighborhood-average student background, all columns expect (8) control for a cubic in the socioeconomic background index. The results in the lower panel control for old-neighborhood fixed effects. Cluster (old-neighborhood)-robust standard errors in parentheses. Significant at * 10%, ** 5%

In Table A7 in the Appendix I present results similar to Table 4, but with VA constructed from teacher grades. As in the previous sub-section, persistent teacher grade VA is more strongly related to teacher grades and less strongly to exam scores than persistent exam score VA. While transitory teacher grade VA predicts exam score, it only weakly predicts teacher grades, similar to exam score VA. As in the previous sub-section, this likely reflects relative grading. However, both persistent and transitory teacher grade VA are about as strongly related to longer-term outcomes as the corresponding exam score indicators.

5.3 Changes in catchment areas

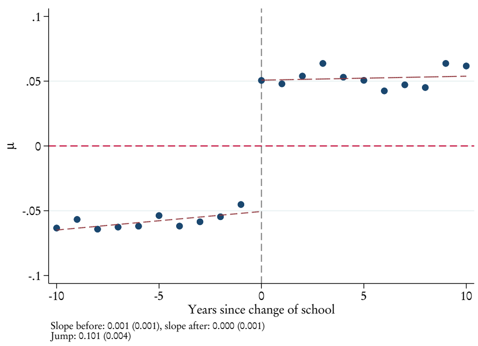
A potential concern with the previous two quasi-experiments is that they are based on students moving. While the analysis shows no indication of unobservable sorting and the context suggests that sorting based on value-added is unlikely, moving students may do so in a way that creates a correlation between value-added and unobserved characteristics of the moving students. In this final quasi-experiment I will study changes in the schools' catchment areas, which arguably are exogenous to the students. As very limited data exist on school catchment areas, I will infer these from the students' neighborhoods, as in the previous subsection. To find neighborhoods that change school assignment, I will identify neighborhoods whose students in each year before some year t overwhelmingly attend one school (meaning that the at least 80 percent of the students in the neighborhoods attend the school, only considering neighborhoods by years with at least four students) and then in t and all following years attend some different school. In analog to the quasi-experiments in the previous sections, I will estimate value-added from the students in neighborhoods that do not change school assignment, and study whether these VA indicators predict outcomes of students in the neighborhoods changing schools, conditional on neighborhoods characteristics or fixed effects. I identify 1,218 neighborhoods that change schools, with a total of 68,466 students. Figures A9 and A10 in the Appendix shows the student-weighted distributions of the years of change and the difference between graduation year and the year of the change.

A challenge interpreting the results from these analyses is that I don't observe the process leading up to and following the school change. I only observe that students from a given neighborhood complete one school before and another after a given year. This can reflect rezoning of existing schools (some neighborhood are transferred from one school to another, e.g. because of imbalances in capacity utilization) or changes in school structure (schools are closed down or new schools opened). Also, as I only observe the school where the students eventually complete compulsory schooling, I do not know for how long students have been attending that school. For students a few years after their neighborhood changed school assignment I don't know whether or for how long they attended the old school before going to the new school. Finally, I do not know the reason for any change. However, as the change is permanent, it seems unlikely to be driven by individual students. Still, as the circumstances concerning the change in catchment areas are unclear, I will disregard the first cohort completing compulsory school at the new school.

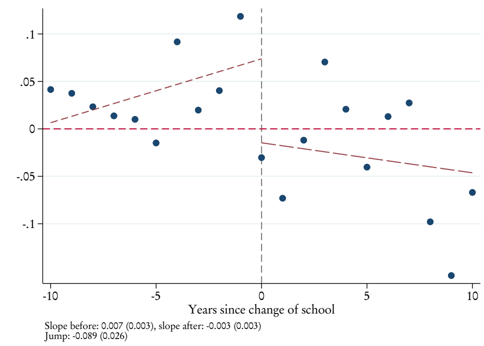
As I can follow neighborhoods and see how students' outcomes evolve over time, this natural experiment lends itself to an event study. In Figure 3 I show outcomes of students in neighborhoods with an absolute change in predicted VA of at least .05 SD. Sub-figure (a) shows the average change in predicted VA. In all the sub-figures of Figure 3 outcomes are multiplied with the sign of the change in predicted VA, such that outcomes are expected to change from on average negative to positive. This is very clear for average predicted VA, which changes from -.05 to .05, i.e. an average absolute change of .1 SD. Except for around the change from old to new school there is little evidence of trends in VA. Sub-figure (b) shows the change in average transitory VA, which changes in the opposite direction of persistent VA.

In sub-figure (c) I show a similar event study using average exam scores. Average exam scores change by .014 SD, in the same direction as the change in predicted VA, but the change is not significant. Finally, sub-figure (c) shows the event study for residualized exam scores, constructed by adjusting for individual student background and transitory VA (estimated from students in units that never change school, like persistent VA). This substantially reduces the dispersion of the yearly averages. Residualized exam scores have an average change of .12 SD. This change is significantly different from zero and not significantly different from the change in predicted VA.

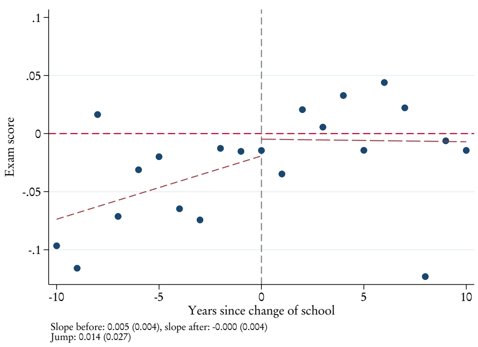
In Table 5 I study the relationship between exam scores of students in neighborhood that change school assignment and value-added estimated from students in never-changing units in a more parametric way, and include students in units whose predicted VA change by less than



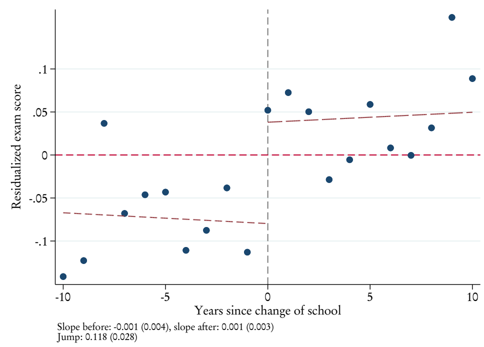
(a) Predicted persistent VA



(b) Predicted transitory VA



(c) Exam score



(d) Residualized exam score

Figure 3: Average absolute change in outcomes following changes in assigned school - event study

Note: Sample is 38,759 studentsgraduating within 10 years a change of predicted school that give $|\Delta\hat{\mu}| > .05$. All outcomes are multiplied with $sign(\Delta\hat{\mu})$, such that values are expected to change from negative to positive. E.g., predicted school quality is in sub-figure (a) is $\tilde{\mu} = \hat{\mu} \cdot sign(\Delta\hat{\mu})$. Subfigure (b) shows observed exam scores, while subfigure (c) shows exam scores residualized by adjusting for student characteristics (X) and transitory VA of the graduating cohort (η). Lines and notes show separate student-level linear fits before and after the change

.05.

Column (1) shows how exam scores are related to VA of the old and new school for students graduating from the new school (in the upper panel) and from the old school (lower panel). Exam scores of students graduating from the new school are significantly related to the persistent VA of both the new and the old school. The relationship is strongest for the new school, where it is not significantly different from one. This is what we would expect if students on average have spent a substantial amount of time in the new school, but also, earlier, in the old school. Exam scores are strongly related to transitory VA of the new school, and unrelated to the transitory VA of the old school. In contrast, exam scores of students graduating from the old school are strongly related to transitory and persistent VA of the old school, and unrelated to the VA of the new school.

The results for teacher grades, in column (2), mostly reflect those for exam scores, although the coefficients are smaller. The most striking difference is that teacher grades are unrelated to transitory VA for graduates from the new school, and only weakly related to transitory VA for graduates from the old school, likely reflecting relative grading. Completion of grade 11 and high school is related to VA of the old school both for graduates from the old and new school. Completion of high school is related to VA also of the new school for both groups of graduates. This may reflect that changes related to the change of school assignment also are concurrent with other changes, which impact students in upper secondary school. However, while completion is related to transitory VA of the old school graduates from the old school, this is not the case for graduates from the new school.[12]

In Table A8 in the Appendix I repeat the analyses in Table 5 with controls for neighborhood fixed effects. This makes the very clear how exam scores are related to the persistent and transitory VA of the new (old) school for students that graduate from the new (school), with little cross effects from the other school. Teacher grades are significantly related to persistent VA of the new school for students graduating from this school, and otherwise not related to VA. The results for longer-term outcomes are mostly to imprecise to be informative.

[12] The sample for earnings consists of fewer cohorts than the other outcomes. Furthermore, students graduating from the new schools will never be in the first cohort and be rare in the next cohorts, few students with earnings data is observed graduating from the new schools. Thus, while earnings of students graduating from the old school is related to VA of the old school, as expected, the results for students graduating from the new school are too imprecise to be informative.

Table 5: Effect of change exam VA from change in predicted school

	(1) Exam score	(2) Teacher grade	(3) Complete year 11	(4) Complete high school	(5) NEET	(6) Background index
Students graduating from new school						
$\hat{\mu}^{New}$	0.866**	0.484**	0.021	0.111*	0.087	0.061
	(0.097)	(0.098)	(0.041)	(0.058)	(0.053)	(0.065)
$\hat{\mu}^{Old}$	0.501**	0.344**	0.153**	0.198**	-0.134**	0.231**
	(0.112)	(0.121)	(0.045)	(0.070)	(0.056)	(0.069)
$\hat{\eta}^{New}$	0.165**	0.012	0.008	0.010	0.001	0.026**
	(0.023)	(0.022)	(0.009)	(0.012)	(0.013)	(0.010)
$\hat{\eta}^{Old}$	0.002	0.029	0.013	-0.000	-0.004	0.013
	(0.050)	(0.059)	(0.015)	(0.029)	(0.023)	(0.024)
Students graduating from old school						
$\hat{\mu}^{New}$	0.078	0.226*	-0.018	0.138**	-0.004	-0.105
	(0.096)	(0.116)	(0.048)	(0.053)	(0.035)	(0.069)
$\hat{\mu}^{Old}$	0.818**	0.341**	0.154**	0.162**	-0.016	0.067
	(0.099)	(0.120)	(0.049)	(0.054)	(0.032)	(0.059)
$\hat{\eta}^{New}$	0.030	0.056	-0.017	0.017	-0.010	-0.015
	(0.051)	(0.054)	(0.020)	(0.023)	(0.014)	(0.022)
$\hat{\eta}^{Old}$	0.270**	0.046**	0.032**	0.022**	-0.009	-0.030**
	(0.021)	(0.021)	(0.010)	(0.010)	(0.007)	(0.010)
N students	66773	66144	61034	48017	37830	66773
Nclusters	1212	1212	1212	1201	1171	1212

Changers live in a neighborhood (basic statistical unit) that changes assigned school. All regressions control for socio-ec index (except (6)), year dummies and dummy before/after. Significant at * 10%, ** 5%

6 Conclusion

In this paper I study school quality in Norwegian compulsory school. In line with previous studies, I find that by controlling exam performance for previous test scores it is possible to construct indicators of persistent school quality that on average show no bias. However, I also find that indicators controlling for socioeconomic variables and not test scores are highly correlated with indicators controlling for previous test scores, and that these indicators are also able to predict exam performance out of sample with little sign of bias. Both types of indicators predict the students' later outcomes, including schooling completed, inactivity and earnings, demonstrating that the differences in exam performance capture important differences in student skills, and that investments in these skills have long-term gains.

Associations between school value-added and student outcomes may reflect an effect of school quality or sorting of students to schools. To address concerns related to student sorting I draw on variation from three quasi-experiments, where students move/change school or the link between residence and school is changed. This allows me to estimate value-added from a group of stayers, and investigate how outcomes of movers depend on the school they attend. In all three settings I find that a change in school value-added is associated with a similar change in exam results. Furthermore, in none of the analyses there is any indication that the identifying assumption, that changes in value-added are conditionally independent of student characteristics, is not satisfied. I thus conclude that the persistent value-added measures are good measures of school quality.

The quasi-experiments also allows me to study how transitory unexplained changes in the stayers' results impact on movers' outcomes. These changes are predictive of movers' outcomes, although with substantial bias. This indicates that while such changes partly reflect unmeasured differences in students' backgrounds, they also measure an important transitory component of school quality. Given what we know about the Norwegian context, this may reflect e.g. differences between the different teachers teaching successive cohorts within the same school.

Taken together, the results underline the importance of school quality for short- and long-term student outcomes. Furthermore, the results point to the relevance and limited bias in indicators controlling either for previous test scores or only for socioeconomic background. This latter set of indicators may be useful as a measure of school quality in school systems with limited early testing (as in Norway), and also allows estimating school quality at early stages of primary school, where prior tests are usually not available, and to study long-term outcomes of students for whom early test data is not available.

Compared to the indicators based on exam grades, indicators based on teacher grades are much less informative about outcomes other than teacher grades. This indicates that while teacher grades are highly predictive at the student level, there are systematic school-level biases in teacher grades, e.g. differences in local grading standards, that make teacher grades less useful for evaluating school quality.

Finally, the analyses quantitatively link school outcomes and quality with students' long-term outcomes. The quasi-experiments do not always allow clear conclusions on the effects of school quality on post-schooling outcomes, but when they do, the results indicate that school quality has important long-term effects. As a large number of studies evaluates different initiatives and policies, this valuation of school quality is important to better interpret the findings from such studies and prioritize resources.

References

Joshua Angrist, Peter Hull, Parag A Pathak, and Christopher R Walters. Simple and credible value-added estimation using centralized school assignment. Working Paper 28241, National Bureau of Economic Research, December 2020. URL http://www.nber.org/papers/w28241.

Joshua D Angrist, Peter D Hull, Parag A Pathak, and Christopher R Walters. Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, 132(2):871–919, 2017.

Manudeep Bhuller, Magne Mogstad, and Kjell G Salvanes. Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4):993–1030, 2017.

Pedro Carneiro, Italo Lï¿œpez Garcï¿œa, Kjell G. Salvanes, and Emma Tominey. Intergenerational mobility and the timing of parental income. *Journal of Political Economy*, 129(3): 757–788, 2021. doi: 10.1086/712443. URL https://doi.org/10.1086/712443.

Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. How does your kindergarten classroom affect your earnings? evidence from project star *. *The Quarterly Journal of Economics*, 126(4):1593–1660, 2011. doi: 10.1093/qje/qjr041. URL + http://dx.doi.org/10.1093/qje/qjr041.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104 (9):2633–79, 2014a.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632, September 2014b. doi: 10.1257/aer.104.9.2593. URL http://www.aeaweb.org/articles?id=10.1257/aer.104.9.2593.

David J Deming, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. School choice, school quality, and postsecondary attainment. *American Economic Review*, 104(3):991–1013, 2014.

Rachel Ekren. Dokumentasjon-beregning av value addedindikatorer for oslo kommune. Technical report, 2015.

Torberg Falch, Ole Henning Nyhus, and Bjarne StrÃžm. Performance of Young Adults: The Importance of Different Skills. *CESifo Economic Studies*, 60(2):435–462, 02 2014. ISSN 1610-241X. doi: 10.1093/cesifo/ifu005. URL https://doi.org/10.1093/cesifo/ifu005.

Jon H Fiva and Lars J Kirkebøen. Information shocks and the dynamics of the housing market. *The Scandinavian Journal of Economics*, 113(3):525–552, 2011.

Peter Fredriksson, Björn Öckert, and Hessel Oosterbeek. Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1):249–285, 2012.

James Heckman and Pedro Carneiro. Human capital policy. Working Paper 9495, National Bureau of Economic Research, February 2003. URL http://www.nber.org/papers/w9495.

C. Kirabo Jackson. What do test scores miss? the importance of teacher effects on nonï¿œtest score outcomes. *Journal of Political Economy*, 126(5):2072–2107, 2018. doi: 10.1086/699018. URL https://doi.org/10.1086/699018.

C. Kirabo Jackson, Shanette C Porter, John Q Easton, and Sebastiï¿œn Kiguel. Who benefits from attending effective schools? examining heterogeneity in high school impacts. Working Paper 28194, National Bureau of Economic Research, December 2020. URL http://www.nber.org/papers/w28194.

Thomas J Kane and Douglas O Staiger. Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research, 2008.

Lars J Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

Jr. Roland G. Fryer. *The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments*, volume 2, pages 95–322. North-Holland, Amsterdam, 2017.

Kjartan Steffensen, Rachel Ekren, Oda O Zachrisen, and Lars J Kirkebøen. Er det forskjeller i skolers og kommuners bidrag til elevenes læring i grunnskolen? en kvantitativ studie. Technical report, 2017.

Appendix

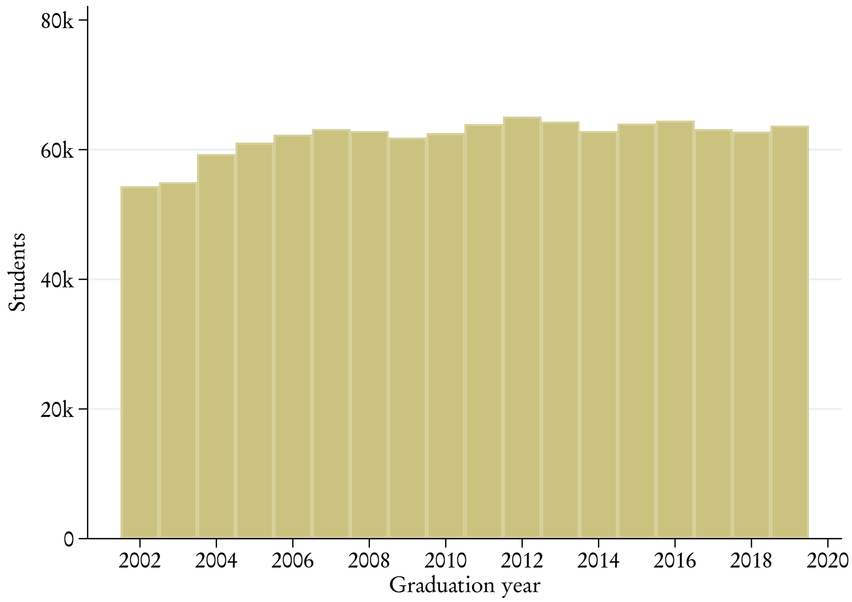


Figure A1: Cohort sizes

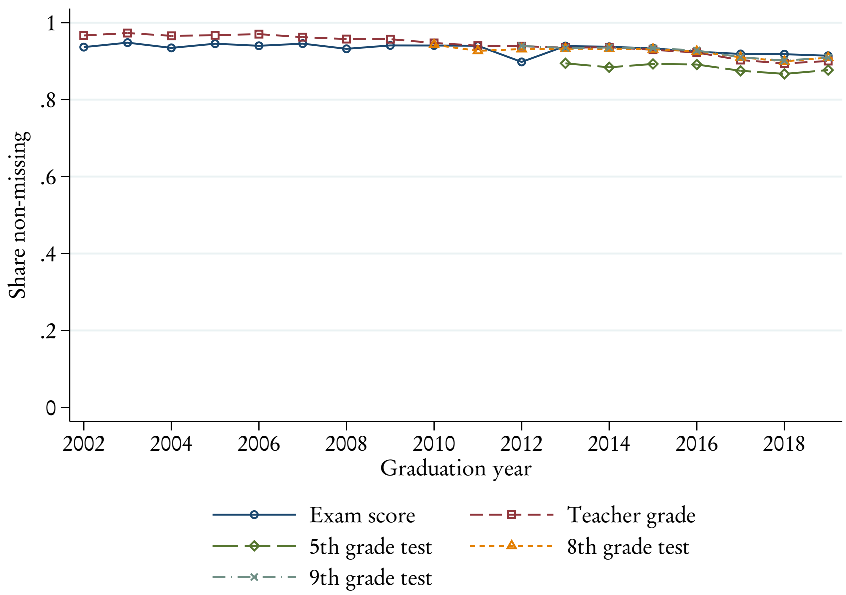


Figure A2: Share of students with non-missing exam scores, teacher grades and test scores by cohort

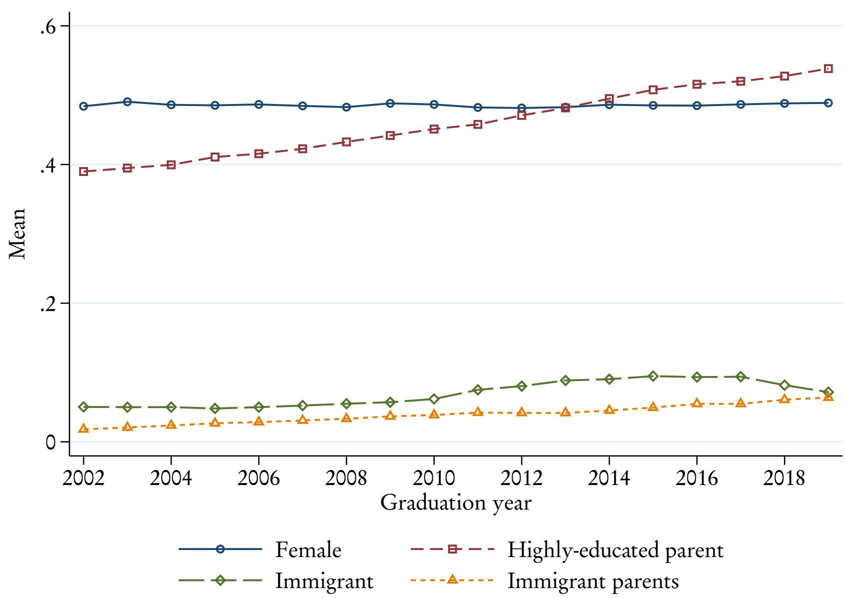


Figure A3: Mean background characteristics by cohort

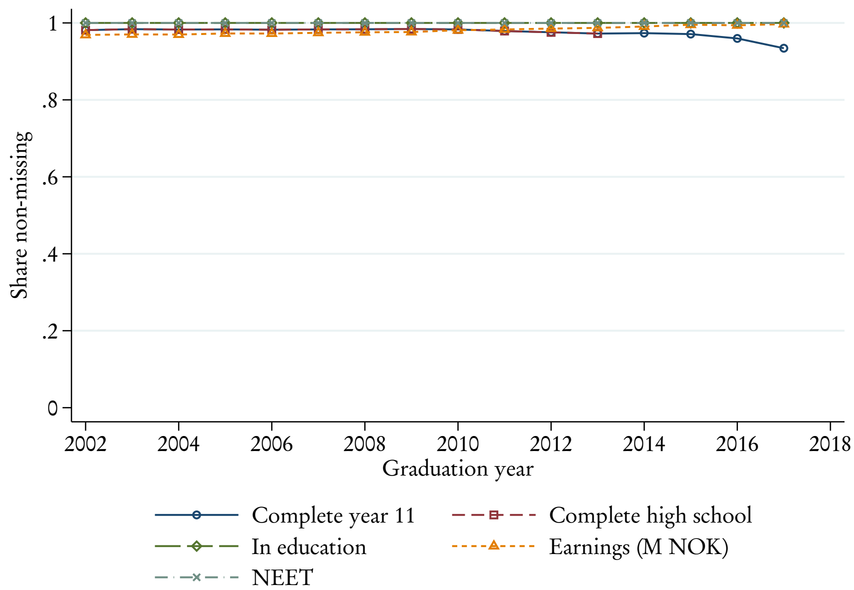


Figure A4: Share of students with non-missing longer-term outcomes by cohort

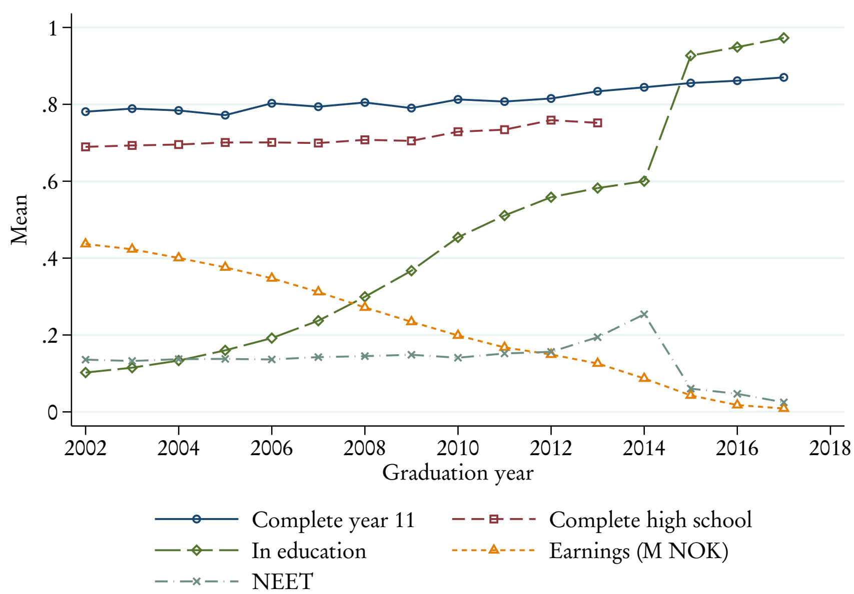


Figure A5: Mean longer-term outcomes by cohort

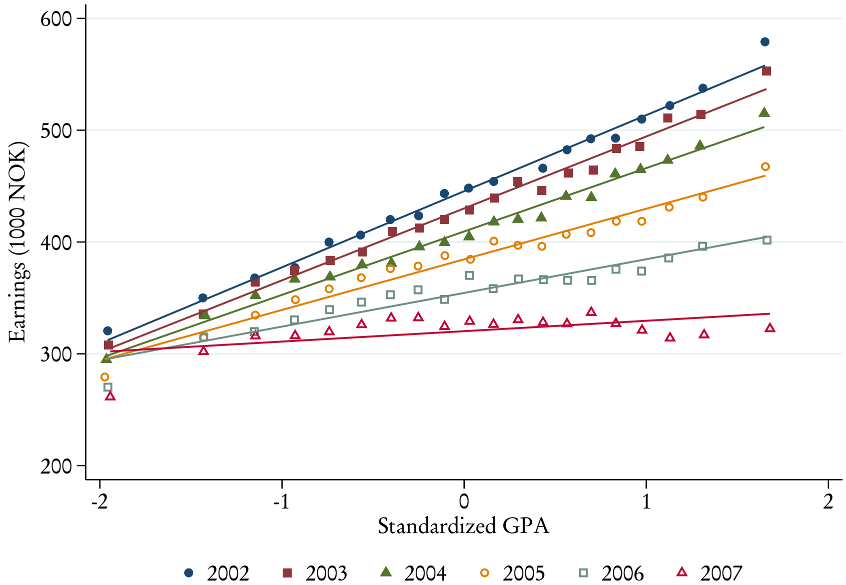


Figure A6: Mean earnings by standardized GPA and cohort

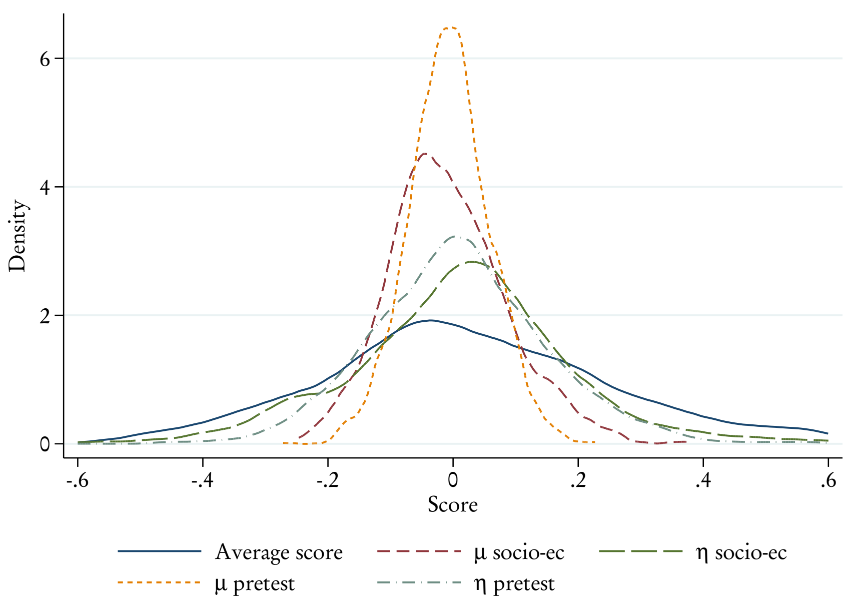


Figure A7: School average exam score and estimated school quality for the 2014 cohort

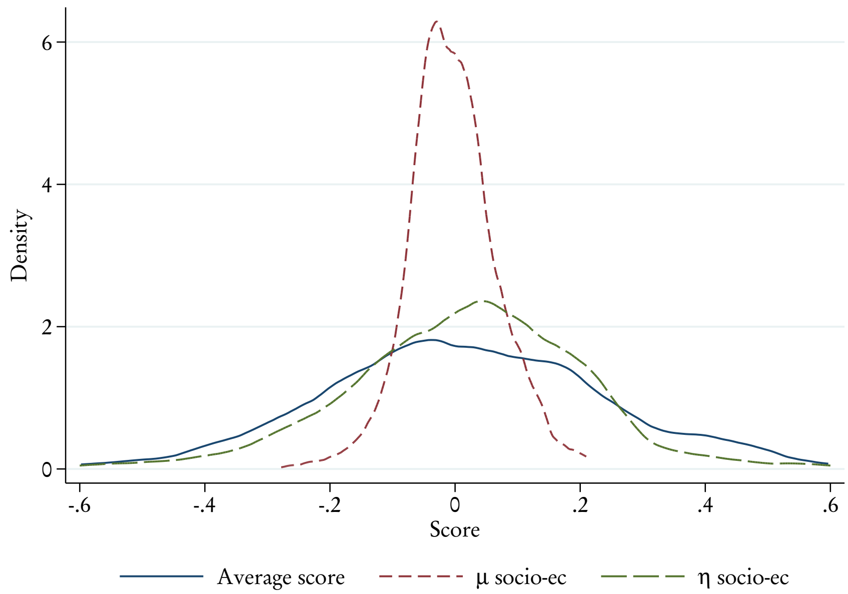


Figure A8: School average exam score and estimated school quality for the 2014 cohort

Table A1: Predicting exam performance of 2019 graduates

	(1)	(2)	(3)	(4)	(5)	(6)
	Indicators control for previous score			Indicators control for background only		
Exam value-added	1.212**	1.064**		1.126**		
	(0.132)	(0.109)		(0.079)		
8th grade value-added	1.000**		0.815**		0.661**	
	(0.106)		(0.096)		(0.063)	
5th grade value-added	0.360**					0.360**
	(0.066)					(0.069)
Controls (cubic + school mean):						
Grade 8 score	No	Yes	No	No	No	No
Grade 5 score	No	No	Yes	No	No	No
Family background	Yes	Yes	Yes	Yes	Yes	Yes
N students	48544	48544	48544	48544	48544	48544
N clusters	1023	1023	1023	1023	1023	1023
R^2	0.146	0.449	0.362	0.146	0.142	0.137

Value-added indicators are constructed from the 2010-2014 cohorts. All regressions control for cubic in an index of socioeconomic background as well as the school mean index (same index as used for constructing indicators) and year dummies. Specifications controlling for cubic in pretest also control for school-mean pretest. Standard error are clustered at the school. Significant at * 10%, ** 5%

Table A2: School quality and short- and long-term outcomes, with municipality-fixed effects

	(1) Written exam	(2) Oral exam	(3) Teacher grade	(4) Completed year 11	(5) Completed high school	(6) Years schooling	(7) In education	(8) Employed	(9) NEET	(10) Earnings	(11) Log earnings
$\hat{\mu}$ written exam	0.729**	0.489**	0.519**	0.174**	0.154**	0.839**	-0.005	0.061**	-0.034	0.896**	0.144**
	(0.135)	(0.083)	(0.106)	(0.034)	(0.034)	(0.289)	(0.020)	(0.024)	(0.024)	(0.183)	(0.033)
$\hat{\mu}$ oral exam	0.519**	0.725**	0.709**	0.111**	0.084**	0.425	-0.025	0.040**	-0.026	0.505**	0.061*
	(0.091)	(0.085)	(0.110)	(0.031)	(0.029)	(0.272)	(0.020)	(0.020)	(0.019)	(0.184)	(0.032)
$\hat{\mu}$ teacher	0.276**	0.384**	0.919**	0.048**	0.067**	0.608**	-0.014	0.023	-0.022	0.113	0.018
	(0.061)	(0.074)	(0.072)	(0.019)	(0.024)	(0.230)	(0.015)	(0.016)	(0.014)	(0.126)	(0.025)
Written exam score	1.000	0.501**	0.621**	0.114**	0.142**	0.896**	0.015**	0.032**	-0.031**	0.522**	0.079**
	(.)	(0.005)	(0.006)	(0.002)	(0.003)	(0.018)	(0.001)	(0.002)	(0.002)	(0.013)	(0.002)
Oral exam score	0.485**	1.000	0.587**	0.109**	0.143**	0.848**	0.013**	0.034**	-0.033**	0.559**	0.085**
	(0.005)	(.)	(0.006)	(0.003)	(0.003)	(0.018)	(0.001)	(0.002)	(0.002)	(0.012)	(0.002)
Teacher grade	0.718**	0.701**	1.000	0.181**	0.233**	1.319**	0.015**	0.059**	-0.055**	0.778**	0.114**
	(0.004)	(0.003)	(.)	(0.003)	(0.002)	(0.017)	(0.002)	(0.002)	(0.002)	(0.018)	(0.003)
N	83371	83371	83144	82871	82871	81317	83371	83371	83371	81297	71667
# clusters	428	428	428	428	428	428	428	428	428	428	428
$\bar{y}$	0.015	0.015	0.037	0.808	0.714	13.784	0.110	0.860	0.114	4.407	1.512

Note: Each cell is a separate regression of outcome on VA indicator or exam/teacher grade on the 2002 and 2003 compulsory school graduation cohorts. Outcomes (1)-(3) are from the end of compulsory school, (4) is observed one year after completing compulsory school and (5) five years after. Outcomes (6)-(12) are observed in 2017, i.e. 14-15 years after graduation from compulsory school, around age 30. (6) is nominal duration of highest completed degree (in years, including compulsory school); (7) is a dummy for whether the person in education in 2017; (8) is an earnings-based employment measure (earnings $>$ G, approx USD 10 000); (9) is a dummy for not in employment, education or training; (10) is annual labor earnings and (11) is log annual earnings. The indicators are constructed from the 2004-2008 cohorts. All regressions control for cubic index of socioeconomic background (same as used in indicators), school*year mean index and year dummies and municipality-fixed effects. Standard errors are clustered at municipality level. Significant at * 10%, ** 5%

Table A3: Predicting exam performance of 2013 graduates

	(1)	(2)	(3)	(4)	(5)	(6)
	Indicators control for previous score			Indicators control for background only		
Exam value-added	1.079**	1.016**		1.018**		
	(0.111)	(0.102)		(0.082)		
8th grade value-added	0.627**		0.670**		0.664**	
	(0.101)		(0.103)		(0.073)	
5th grade value-added	0.399**					0.346**
	(0.057)					(0.061)
Controls (cubic + school mean):						
Grade 8 score	No	Yes	No	No	No	No
Grade 5 score	No	No	Yes	No	No	No
Family background	Yes	Yes	Yes	Yes	Yes	Yes
N students	46950	46953	46953	46953	46953	46950
N clusters	1013	1013	1013	1013	1013	1013
R^2	0.228	0.481	0.424	0.228	0.224	0.221

Value-added indicators are constructed from the 2015-2017 cohorts. All regressions control for cubic in an index of socioeconomic background as well as the school mean index (same index as used for constructing indicators) and year dummies. Specifications controlling for cubic in pretest also control for school-mean pretest. Standard error are clustered at the school. Significant at * 10%, ** 5%

Table A4: Predicting completion of upper secondary for 2013 graduates

	(1)	(2)	(3)	(4)	(5)	(6)
	Indicators control for previous score			Indicators control for background only		
Exam value-added	0.134**	0.104**		0.176**		
	(0.044)	(0.042)		(0.031)		
8th grade value-added	0.116**		0.096**		0.142**	
	(0.037)		(0.034)		(0.025)	
5th grade value-added	0.106**					0.098**
	(0.021)					(0.021)
Controls (cubic + school mean):						
Grade 8 score	No	Yes	No	No	No	No
Grade 5 score	No	No	Yes	No	No	No
Family background	Yes	Yes	Yes	Yes	Yes	Yes
N students	48005	48008	48008	48008	48008	48005
N clusters	1018	1018	1018	1018	1018	1018
R^2	0.045	0.118	0.090	0.045	0.045	0.045

Value-added indicators are constructed from the 2015-2017 cohorts. All regressions control for cubic in an index of socioeconomic background as well as the school mean index (same index as used for constructing indicators) and year dummies. Specifications controlling for cubic in pretest also control for school-mean pretest. Standard error are clustered at the school. Significant at * 10%, ** 5%

Table A5: Exam score VA and outcomes of shool-changers, school fixed effects

	(1) Written exam	(2) Teacher grade	(3) Completed year 11	(4) Completed high school	(5) NEET	(6) Control (pretest/ index)
Indicators controlling for pretest						
$\hat{\mu}$ year 8-school	0.121	-0.749	-0.628	0.110	-1.219	-0.988
	(1.606)	(2.290)	(2.723)	(1.651)	(2.834)	(1.467)
$\hat{\mu}$ year 9-school	0.217**	0.818**	0.460**	0.090	-0.081	-0.073
	(0.095)	(0.142)	(0.151)	(0.091)	(0.189)	(0.077)
$\hat{\eta}$ year 8-school	0.024	0.065	0.016	-0.005	-0.020	-0.101
	(0.130)	(0.192)	(0.229)	(0.138)	(0.245)	(0.121)
$\hat{\eta}$ year 9-school	0.055*	0.291**	-0.017	-0.032	-0.027	0.024
	(0.031)	(0.052)	(0.047)	(0.032)	(0.064)	(0.028)
Indicators controlling for family background						
$\hat{\mu}$ year 8-school	2.860	1.164	0.688	1.115	0.575	0.037
	(1.814)	(1.725)	(1.988)	(1.248)	(2.438)	(0.984)
$\hat{\mu}$ year 9-school	0.440**	0.724**	0.463**	0.046	-0.050	-0.026
	(0.102)	(0.105)	(0.115)	(0.067)	(0.129)	(0.054)
$\hat{\eta}$ year 8-school	0.306*	0.223	0.123	0.090	0.196	-0.014
	(0.172)	(0.171)	(0.196)	(0.118)	(0.247)	(0.094)
$\hat{\eta}$ year 9-school	0.161**	0.302**	0.074	0.001	-0.011	-0.001
	(0.049)	(0.050)	(0.052)	(0.029)	(0.073)	(0.024)
Student controls	Yes	Yes	Yes	Yes	Yes	*
School controls	Yes	Yes	Yes	Yes	Yes	Yes
N students	7874	8014	7828	6092	2112	6160
N clusters	935	935	932	904	692	906

See notes to Table 3. Significant at * 10%, ** 5%

Table A6: Teacher grade VA and outcomes of school-changers

	(1)	(2)	(3)	(4)	(5)	(6)
	Written exam	Teacher grade	Completed year 11	Completed high school	NEET	Control (pretest/ index)
Indicators controlling for pretest						
$\hat{\mu}$ year 8-school	-0.027	0.018	0.062	-0.001	0.131	0.083*
	(0.051)	(0.088)	(0.081)	(0.050)	(0.094)	(0.045)
$\hat{\mu}$ year 9-school	0.035	-0.029	0.934**	-0.096*	-0.013	0.030
	(0.053)	(0.084)	(0.088)	(0.050)	(0.098)	(0.048)
$\hat{\eta}$ year 8-school	0.042	0.132**	0.159**	0.040	0.056	-0.039
	(0.032)	(0.053)	(0.054)	(0.031)	(0.058)	(0.030)
$\hat{\eta}$ year 9-school	0.037	-0.042	0.471**	0.002	0.064	0.011
	(0.033)	(0.051)	(0.051)	(0.032)	(0.056)	(0.029)
Indicators controlling for family background						
$\hat{\mu}$ year 8-school	0.062	0.323**	0.151*	0.189**	0.044	0.159*
	(0.046)	(0.085)	(0.083)	(0.085)	(0.046)	(0.089)
$\hat{\mu}$ year 9-school	0.028	0.418**	0.297**	0.980**	-0.032	0.093
	(0.043)	(0.083)	(0.081)	(0.084)	(0.046)	(0.087)
$\hat{\eta}$ year 8-school	0.034	0.067	0.061	0.118**	0.039	0.052
	(0.028)	(0.047)	(0.049)	(0.053)	(0.028)	(0.052)
$\hat{\eta}$ year 9-school	0.018	0.210**	0.046	0.474**	0.051*	0.131**
	(0.025)	(0.046)	(0.048)	(0.050)	(0.028)	(0.055)
Student controls	Yes	Yes	Yes	Yes	Yes	*
School controls	Yes	Yes	Yes	Yes	Yes	Yes
N students	7874	8014	7828	6092	2112	6160
N clusters	935	935	932	904	692	906

See notes to Table 3. Significant at * 10%, ** 5%

Table A7: Teacher grade-VA and movers' outcomes

	(1) Written exam	(2) Teacher grade	(3) Complete year 11	(4) Complete high school	(5) Earnings	(6) Log earnings	(7) NEET	(8) Background index
$\hat{\mu}^{Old}$	0.166**	0.325**	0.071**	0.114**	0.503**	0.031	-0.065**	0.121**
	(0.028)	(0.030)	(0.013)	(0.017)	(0.235)	(0.037)	(0.016)	(0.018)
$\hat{\mu}^{New}$	0.401**	0.721**	0.076**	0.105**	0.270	0.044	-0.044**	-0.047**
	(0.028)	(0.028)	(0.013)	(0.017)	(0.217)	(0.037)	(0.015)	(0.014)
$\hat{\eta}^{Old}$	0.086**	0.073**	0.022**	0.031**	0.073	-0.004	-0.016**	0.031**
	(0.014)	(0.015)	(0.007)	(0.009)	(0.081)	(0.017)	(0.008)	(0.007)
$\hat{\eta}^{New}$	0.390**	0.072**	0.044**	0.064**	0.268**	0.050**	-0.016*	0.003
	(0.014)	(0.015)	(0.007)	(0.009)	(0.079)	(0.017)	(0.008)	(0.007)
With neighborhood fixed effects:								
$\hat{\mu}^{Old}$	0.038	0.024	0.000	0.037	1.001	-0.013	-0.003	0.035
	(0.064)	(0.066)	(0.031)	(0.042)	(0.826)	(0.178)	(0.049)	(0.030)
$\hat{\mu}^{New}$	0.342**	0.636**	0.064**	0.081**	-0.038	0.065	-0.041**	-0.053**
	(0.031)	(0.031)	(0.014)	(0.018)	(0.324)	(0.050)	(0.018)	(0.015)
$\hat{\eta}^{Old}$	0.036**	0.029*	0.003	-0.001	0.091	-0.058*	-0.010	0.025**
	(0.017)	(0.017)	(0.008)	(0.011)	(0.143)	(0.031)	(0.011)	(0.008)
$\hat{\eta}^{New}$	0.386**	0.066**	0.042**	0.063**	0.259**	0.054**	-0.019**	0.008
	(0.015)	(0.015)	(0.008)	(0.009)	(0.100)	(0.022)	(0.009)	(0.007)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
N students	95806	98228	94270	71190	18142	14593	52794	104656
N clusters	10441	10477	10387	9793	6182	5582	9005	10587

Note: Sample is students moving during compulsory school. Outcomes are the same as in Table 3.$\hat{\mu}^{Old}$ and $\hat{\eta}^{Old}$ are persistent and transitory VA (teacher grades adjusted for student background) of the modal school lower secondary school of the student's neighborhood when starting school and $\hat{\mu}^{New}$ and $\hat{\eta}^{New}$ are similar VA of the lower secondary school of the student's neighborhood after moving. Cluster (neighborhood)-robust standard errors in parentheses. Significant at * 10%, ** 5%

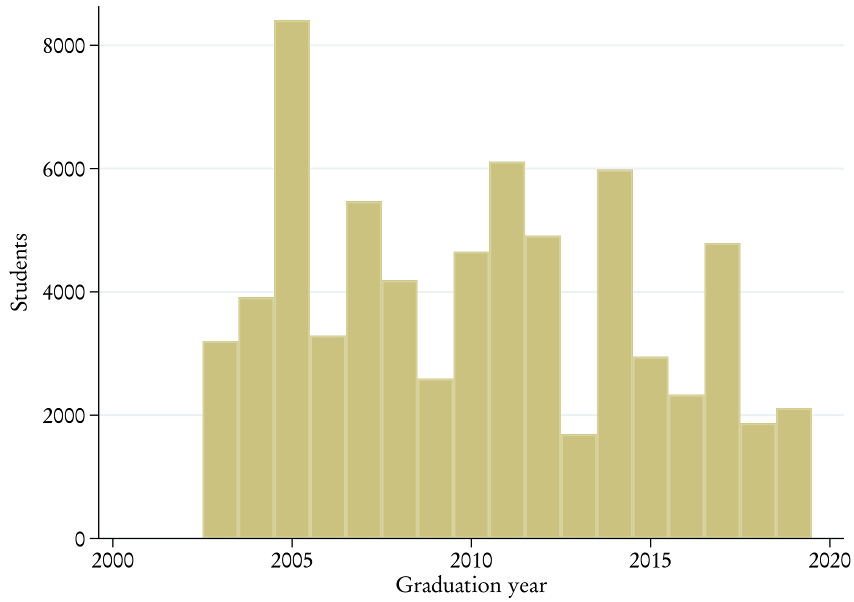


Figure A9: Year of change predicted school

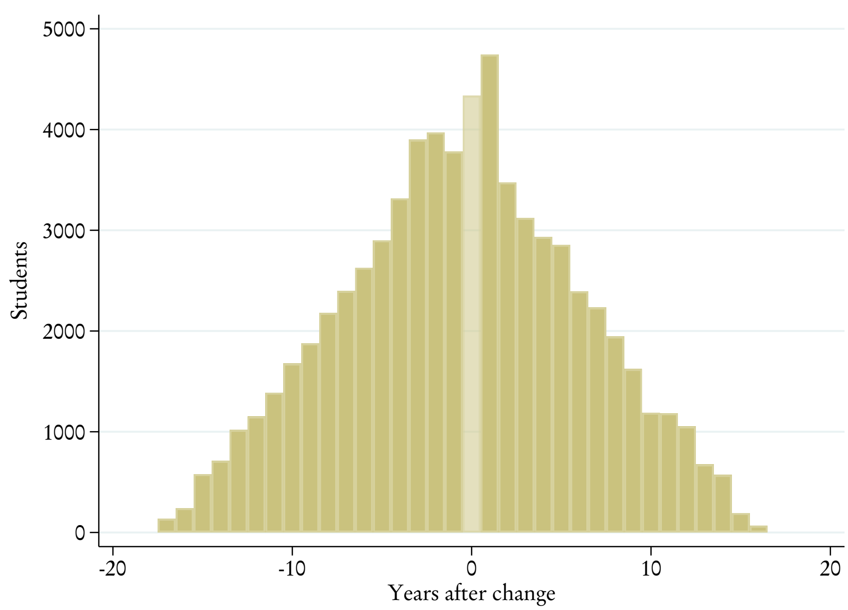


Figure A10: Years since change predicted school

Table A8: Effect of change in assigned school quality on exam scores, neighborhood fixed effects

	(1) Exam score	(2) Teacher grade	(3) Complete year 11	(4) Complete high school	(5) NEET	(6) Background index
Students graduating from new school						
$\hat{\mu}^{New}$	0.899**	0.635**	0.056	0.234	0.119	-0.120
	(0.286)	(0.274)	(0.112)	(0.171)	(0.155)	(0.130)
$\hat{\mu}^{Old}$	0.360	-0.082	0.059	-0.041	-0.039	0.306**
	(0.275)	(0.293)	(0.113)	(0.146)	(0.153)	(0.136)
$\hat{\eta}^{New}$	0.170**	0.005	0.010	0.024*	-0.002	0.010
	(0.026)	(0.023)	(0.009)	(0.013)	(0.015)	(0.010)
$\hat{\eta}^{Old}$	0.027	0.023	0.012	-0.019	-0.001	0.019
	(0.056)	(0.055)	(0.016)	(0.029)	(0.027)	(0.023)
Students graduating from old school						
$\hat{\mu}^{New}$	0.103	0.290	0.044	0.197	-0.137	-0.117
	(0.320)	(0.298)	(0.127)	(0.189)	(0.166)	(0.141)
$\hat{\mu}^{Old}$	0.804**	-0.122	0.161	0.048	0.104	0.146
	(0.265)	(0.285)	(0.109)	(0.135)	(0.136)	(0.129)
$\hat{\eta}^{New}$	0.069	0.022	-0.020	0.015	-0.014	-0.003
	(0.058)	(0.055)	(0.024)	(0.027)	(0.018)	(0.023)
$\hat{\eta}^{Old}$	0.268**	0.024	0.033**	0.016	0.007	-0.020
	(0.026)	(0.025)	(0.011)	(0.012)	(0.010)	(0.012)
N students	66773	66144	61034	48017	37830	66773
Nclusters	1212	1212	1212	1201	1171	1212

Changers live in a basic statistical unit that do change assigned school. All regressions control for socio-ec index (except (6)), year dummies and dummy before/after. Cluster (neighborhood)-robust standard errors in parentheses. Significant at * 10%, ** 5%

Table A9: Effect of change in assigned school quality (teacher grades) on exam scores

	(1) Exam score	(2) Teacher grade	(3) Complete year 11	(4) Complete high school	(5) NEET	(6) Background index
Students graduating from new school						
$\hat{\mu}^{New}$	0.899**	0.635**	0.056	0.234	0.119	-0.120
	(0.286)	(0.274)	(0.112)	(0.171)	(0.155)	(0.130)
$\hat{\mu}^{Old}$	0.360	-0.082	0.059	-0.041	-0.039	0.306**
	(0.275)	(0.293)	(0.113)	(0.146)	(0.153)	(0.136)
$\hat{\eta}^{New}$	0.170**	0.005	0.010	0.024*	-0.002	0.010
	(0.026)	(0.023)	(0.009)	(0.013)	(0.015)	(0.010)
$\hat{\eta}^{Old}$	0.027	0.023	0.012	-0.019	-0.001	0.019
	(0.056)	(0.055)	(0.016)	(0.029)	(0.027)	(0.023)
Students graduating from old school						
$\hat{\mu}^{New}$	0.103	0.290	0.044	0.197	-0.137	-0.117
	(0.320)	(0.298)	(0.127)	(0.189)	(0.166)	(0.141)
$\hat{\mu}^{Old}$	0.804**	-0.122	0.161	0.048	0.104	0.146
	(0.265)	(0.285)	(0.109)	(0.135)	(0.136)	(0.129)
$\hat{\eta}^{New}$	0.069	0.022	-0.020	0.015	-0.014	-0.003
	(0.058)	(0.055)	(0.024)	(0.027)	(0.018)	(0.023)
$\hat{\eta}^{Old}$	0.268**	0.024	0.033**	0.016	0.007	-0.020
	(0.026)	(0.025)	(0.011)	(0.012)	(0.010)	(0.012)
N students	66773	66144	61034	48017	37830	66773
Nclusters	1212	1212	1212	1201	1171	1212

Changers live in a basic statistical unit that do change assigned school. All regressions control for socio-ec index (except (6)), year dummies and dummy before/after. Cluster (neighborhood)-robust standard errors in parentheses. Significant at * 10%, ** 5%