Statistisk sentralbyrå
Statistics Norway

School value-added and long-term student outcomes

Lars J. Kirkebøen

TALL SOM FORTELLER

DISCUSSION PAPERS

970

Discussion Papers No. 970, November 2021
Statistics Norway, Research Department

Lars J. Kirkebøen

School value-added and long-term student outcomes

Abstract:
Several recent studies find that interventions in schools can have important lasting consequences for students, and that schools differ in their contribution to students' learning. However, there is less research investigating how these differences between schools influence longer-term outcomes, especially outside the US. In this paper I study school value-added (VA) in Norwegian compulsory school, where between-school differences are smaller than in the US. I find that VA indicators are able to predict in-school performance without bias. Furthermore, VA is strongly related to long-term outcomes, and differences between schools in VA correspond to meaningful differences in long-term outcomes. For example, a one standard deviation higher VA correspond to 1.5 percent higher earnings around age 32. 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. Analysis of teacher grades and exam scores suggest that the former is heavily influenced by relative grading, and that the effect of exam score VA on long-term outcomes reflects the effects of skills acquired in school. In addition to shedding lights on the differences in and mechanisms of school quality, 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 classification: J24, I2

Acknowledgements: I thank Martin Andresen, Nicolai Borgen, Eric Hanushek, Edwin Leuven, Oddbjørn Raaum, Astrid M. J. Sandsør, Ola Vestad, Simon Widerhold and participants in several seminars for valuable feedback and suggestions. The Norwegian Research Council supported this research under projects no. 238050 and 275906.

Address: Lars J. Kirkebøen, Statistics Norway, Research Department. E-mail: kir@ssb.no

Discussion Papers	comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

The Discussion Papers series presents results from ongoing research projects and other research and analysis by SSB staff. The views and conclusions in this document are those of the authors

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no/en/forskning/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

En rekke nyere studier finner at forskjellige tiltak i skolen kan ha viktige varige konsekvenser for elevene, samt at vesentlige forskjeller i skolers bidrag til elevenes læring. Det er imidlertid mindre forskning som undersøker hvordan disse forskjellene mellom skoler påvirker elevenes langsiktige resultater, og det meste av denne forskningen baserer seg på data fra USA.

I denne artikkelen studerer jeg forskjeller i skolebidrag blant norske grunnskoler, der forskjellene i elevsammensetning og læringsutbytte er mye mindre enn mellom skoler i USA. Jeg finner at indikatorer for stabilt skolebidrag i gjennomsnitt gir et presist mål på skoleresultatene til andre elevkull enn de som beregningen baserer seg på. Videre er indikatorene sterkt knyttet til elevenes langsiktige resultater i utdanning og arbeidsmarked etter fullført grunnskole, og forskjeller mellom skoler i bidrag svarer til betydelige langsiktige forskjeller. For eksempel tilsvarer ett standardavvik høyere stabilt skolebidrag 1,5 prosent høyere inntekt rundt 32 år. Forskjellen mellom en skole så vidt blant de 10 prosent med lavest bidrag og en så vidt blant de 10 prosent med høyest bidrag svarer til svarer til en forskjell på 7 prosentpoeng i fullføring av videregående opplæring, og 1-2 prosentpoeng forskjell i sysselsetting.

Tre forskjellige analyser av elever som flytter / bytter skole samt av grunnkretser som endrer skoletilhørighet tyder på at forskjellene i beregnet skolebidrag gjenspeiler forskjeller mellom skolene, og ikke uobserverte forskjeller mellom elever eller nabolag. Separate analyser av standpunkt- og eksamenskarakterer tyder på at standpunktkarakterer er sterkt påvirket av relativ karaktersetting, og at effekten av skolebidrag beregnet fra eksamenskarakterer på langsiktige utfall gjenspeiler effekten av ferdigheter tilegnet i skolen. I tillegg til å øke forståelsen av forskjeller i og mekanismene for skolers bidrag, bidrar funnene i analysene til å koble læringsutbytte med senere arbeidsmarkedsutfall. Dette er verdifullt f.eks. for nytte-kostnadsanalyse av tiltak i skolen.

1 Introduction

Primary and secondary schooling have in several 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 student performance (Roland G. Fryer [2017]) and in identifying effective teachers and schools. There is substantial recent research showing that the contributions to students' learning varies between schools, and can be predicted accurately using value-added (VA) models, and there is active research into how available data best can be used to provide good VA estimates [Deming, 2014, Angrist et al., 2016, 2017]. A smaller literature investigates how differences in school VA matter for long-term student outcomes [Deming et al., 2014].[1] Finally, recent years have seen increasing interest in VA indicators based on non-test outcomes, with Jackson [2018], Jackson et al. [2020] finding that non-test school quality is even more important for longer-term outcomes than schools' effects on test scores. However, there is still limited evidence on the validity of VA models outside the US, and also limited evidence of the long-term consequences of differences in school quality.

In this paper I study VA of Norwegian compulsory schools and how VA relates to long-term student outcomes. I also investigate the validity of the VA estimates as indicators of school quality and some potential mechanisms for the associations with long-term outcomes. I estimate persistent VA using leave-out-year shrinkage estimators where VA for a given year is predicted from other years, similar to the approach Chetty et al. [2014b] use to estimate teacher VA. Detailed population-level administrative data allow me to construct measures of student background that let me estimate credible VA models, observe student outcomes into their early 30s, and track students that change schools.

I find persistent differences in VA between schools. The VA indicators predict the exam scores, teacher grades and longer-term outcomes of students outside the sample used to estimate the VA indicators. For the in-school outcomes I cannot reject that the indicators are forecast-unbiased, as defined by Chetty et al. [2014b]. The relationships between VA and long-term student outcomes are mostly as strong as or stronger than the corresponding cross-sectional student-level relationships. That is, the predicted gain from attending a high-VA school is for most outcomes greater than that associated with a difference in student background corresponding to a similar difference in learning outcomes. Despite Norway being a country with very

[1] There is a closely related literature on teacher VA, investigating both how to estimate VA and long-term effects of high-VA teachers [Kane and Staiger, 2008, Hanushek and Rivkin, 2010, Chetty et al., 2014a,b, Rothstein, 2017]. The teacher VA literature differs from the school VA literature in that it also needs to consider the potential within-school matching of students and teacher based on characteristics observable to the principals, but unobservable to the researcher.

small between-school differences,[2] the differences in VA correspond to meaningful differences in student outcomes, both in school and in the labor market. For example, a one-standard deviation difference in VA corresponds to a 0.5 percentage points difference in labor market participation and a 1.9 percent earnings difference (given positive earnings).

While previous US studies have found that VA estimates controlling for students' previous achievement produce unbiased estimates, lack of test data makes this approach infeasible for older Norwegian cohorts. Instead I use the rich register data to construct measures of family background. The forecast-unbiasedness of the VA indicators shows that adjustment only for contemporaneously observed family background may provide informative VA estimates, at least in some contexts. In addition to allowing estimation of VA in school systems without sufficient data for historical standardized tests, like Norway, this also allows the study of VA also before the first test are available, e.g. for the first years of primary school.

For recent cohorts I also observe standardized tests throughout compulsory school. I briefly lay out how studying different periods and using different sets of controls impact the interpretation and comparison of VA indicators. Investigating VA measures for different stages of compulsory school, I find that these are consistent, and that VA towards the end of compulsory school matters most for exam scores and completion of upper secondary.

My main analysis does not distinguish between a school and the students of that school. Furthermore, most Norwegian students attend their neighborhood school. Thus, while the VA estimator for a given year only depends on outcomes of students of other years, there may be persistent differences between schools in unobserved student characteristics if there are such differences between neighborhoods. To address this I study three different quasi-experiments, where students move or neighborhoods change schools. In each of these I find that the outcomes of movers correspond to what we would expect based on VA estimated from the non-movers, suggesting that the VA indicators reflect school quality, not unobserved characteristics of students or neighborhoods.

Students moving and school closures/openings/rezoning are unlikely to be random, which may be a cause of concern for the validity of the quasi-experiments. However, I find no indication of the movers sorting to high-VA schools based on observed characteristics. This suggests that while moving may be non-random, the VA of movers' new schools are random (conditional on observed characteristics of

[2] OECD [2006], which is roughly contemporaneous with the cohorts studied in the main analyses, report that the between-school variance of student performance in Norway is 6.5 percent of the total variance across all participating countries. The corresponding OECD average is 33.6 percent. The between-school variance explained by a socio-economic index for students and schools is 2.9 percent of the total variance in Norway, while the OECD average is 23.0 percent.

students), and thus that studying the movers constitute a valid quasi-experiment for the effect of school VA on this group.

Studying movers also allows me to investigate transitory VA, that is, year-by-year differences in student performance over and above what can be explained by student characteristics and persistent VA. In Norwegian schools, consecutive cohorts are often taught by different teachers, which may give rise to within-school time variation in VA. The quasi-experiments allow me to study how transitory VA estimated from non-moving students impact the outcomes of incoming students. Transitory VA is strongly associated with outcomes of movers, but, unlike persistent VA, not forecast-unbiased. This suggests that transitory VA reflects in part year-to-year variation in school quality (e.g. in the form of differences between teachers within schools) and in part year-to-year variation in unobserved student characteristics.

In addition to exam score VA I also study VA based on end-of-compulsory teacher grades. While these VA estimates are forecast-unbiased for teacher grades, teacher grade VA is not as strongly associated with long-term outcomes as exam score VA. This is despite long-term outcomes being more strongly associated with teacher grades than with exam scores at the student level. This likely indicates that the classroom teachers are better able to observe students' qualifications throughout the school year than external teachers grading a written exam, but that relative grading practices make school-average teacher grades a worse measure of skill than average exam scores. This interpretation is supported by the quasi-experiments, where I find that moving students do benefit from transitory exam score VA at the incoming school, but don't benefit from transitory teacher grade VA.

The weaker association of teacher grade VA with long-term outcomes is also informative of the mechanisms through which school quality impacts later outcomes. Admission to upper secondary schools is based on GPA from lower secondary, such that higher GPA will provide more educational opportunities. However, teacher grades have approximately 10-20 times the weight of written exam scores in the calculation of GPA.[3] Thus, the finding that exam score-VA indicators matter more than teacher grade-VA inndicators suggest that this mechanism is not very important. Rather, schools contribute to later outcomes by providing skills, which, because of differences in grading practices, are better measured at the school level by exam scores than by teacher grades.

This paper makes several contributions to the VA literature. First, studying a new setting provides additional evidence that VA estimators can provide valid estimates of school quality and point to important differences between schools, also in a context with smaller between-school differences. Furthermore, this paper demon-

[3]The relative weight varies between years, with smaller weight on exams in earlier years, and to a smaller extent between students within years.

strates 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 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. The significant but smaller than one-to-one relationship between transitory VA and entrants' outcomes strongly suggests that transitory VA reflects both within-school differences between cohorts in unobserved characteristics and school VA. Furthermore, from the dispersion of estimated transitory VA and the relationship between transitory VA and the outcomes of an entrant, we can conclude that instruction quality both have a substantial persistent school-level component, and a more volatile component. The latter 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 what 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. 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 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. In Section 5 I present and compare different VA measures, and in Section 6 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). The school system is almost exclusively public, with less than 5 percent of compulsory school students attending private schools.[4] 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 is subject to capacity.

Norwegian schools generally don't have grade teachers. Instead, 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

[4]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 represent a faith-based or pedagogical alternative to the public schools. Less than than 0,5 percent of students attend international schools not funded by the government.

which they teach students from different classes.[5]

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 13 subjects, and sit one oral and one written exam. The average of these grades constitutes the student's grade point average (GPA).

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 in upper secondary education (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 five years.

2.2 Data on student background and outcomes

The data used in this paper are administrative data on standardized tests and end-of-compulsory school grades for the entire student population graduating in the years 2002-2019. Figure A1 in the Appendix shows the number or students per cohort, which mostly varies around 60,000 students. In the following I will index students by their (end of compulsory) graduation year. Thus, while exam scores and teachers grades are available for all graduation cohorts 2002-2019, the 2010 graduation cohort is the first for whom the 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.

[5] 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.

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 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 completion of upper secondary, completed years of schooling and labor earnings. Post-compulsory school outcomes are measured up to or in 2019 (except completion of high school, which is also observed in 2020), i.e. 17 years after the first cohort graduates from compulsory school, and when these students are about 33 years old. As completed education and earnings 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 "on-time" high school completion, which is measured five years after graduation from lower secondary,[6] and thus is observable for cohorts graduating in 2015 or earlier.

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 of student i in school s (z_{is}) are observed at the end of two periods t =0 and 1, corresponding to primary and lower secondary school. Results in primary school z^0_{is} depend on quality in primary school Q^0_s, student characteristics x_{is}, and an idiosyncratic error term:

$$z^0_{is} = Q^0_s + x_{is}\beta^0 + \epsilon^0_{is} \tag{1}$$

Allowing for some persistence in results from primary to lower secondary, captured by the coefficient λ, results in lower secondary school z^1_{is} can be expressed as a function of previous results, school quality in lower secondary, and student background;[7]

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

[7] While students' characteristics are the same at t=0 and 1, the associations with school results

$$
\begin{aligned}
z_{is}^1 &= \lambda z_{is}^0 + Q_s^1 + x_{is}\beta^1 + \epsilon_{is}^1 \\
&= \lambda Q_s^0 + Q_s^1 + x_{is}(\lambda\beta^0 + \beta^1) + (\lambda\epsilon_{is}^0 + \epsilon_{is}^1), \qquad (2)
\end{aligned}
$$

where the second equality makes clear that we can substitute for previous results z_{is}^0 from (1) to express z_1 as a function of school quality in primary and lower secondary and student background characteristics.

I assume that x_{is} capture all sources of student-level persistence in results and Q_s^0 and Q_s^1 all school-level sources, such that the error terms ϵ_{is}^0 and ϵ_{is}^1 are independent with expectation zero, and also uncorrelated with school quality and observed characteristics. With these assumptions, reorganizing (2), the difference between observed results in lower secondary and results expected from the students' background and previous results reflect school quality in lower secondary:

$$Q_s^1 = E_s[z_{is}^1 - \lambda z_{is}^0 - x_{is}\beta^1] \qquad (3)$$

Eq. (3) 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:

$$Q_{av} = \lambda Q_0 + Q_1 = E[z_1 - x(\lambda\beta^0 + \beta^1)] \qquad (4)$$

3.1 Estimating school quality

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

$$z_{ist} = x_{ist}\beta + \epsilon_{ist} \qquad (5)$$

Here, z_{ist} represent the results (typically exam or test scores) of student i graduating from school s at time t.[8]

The vector of covariates (x_{ist}) will always include a cubic in socioeconomic index, defined as $X_{ist} = \tilde{x}_{ist}\hat{\beta}$ for a set of socioeconomic variables $\tilde{x}$, as well as a school-by-cohort average value of this index. To construct the index I regress exam score on a set of dummies for gender*immigration status (native, immigrant, immigrant

may differ.

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

parents)*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.[9] 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 (3) and (4), 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 (3), while controlling only for background characteristics gives a composite measure of quality for both primary and lower secondary, as in (4). 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 (5), I obtain estimated school-by-cohort residuals by taking school-by-cohort averages 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. [2014b] 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.[10] School quality

[9] 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 6, 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.

[10] Data linking students to teachers is not available, such that it is not possible to study teacher

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], Deming et al. [2014], Angrist et al. [2016]), we expect $\hat{\mu}_{st}$ to be reflected in the school results of a 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).

3.2 Evaluating effects of school quality

I will next study associations between estimated persistent VA and short-term (exams, teachers grades) and long-term outcomes (further education, earnings). The general regression equation relating each outcome y_{ist} of a student i graduating from school s at time t on to estimated school quality and student and school characteristics x_{ist} is:

$$y_{ist} = \gamma_1 \hat{\mu}_{st} + \gamma_2 \hat{\eta}_{st} + \theta x_{ist} + \nu_{ist} \tag{6}$$

The controls x_{ist} include a cubic in the socioeconomic index (X_{ist}), school*cohort means of the index, and year dummies, i.e. the same variables as used to estimate VA above. The γ coefficients measure 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 forecasts

VA.

outcomes without error.

$\hat{Q}_{st}$, 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 implies the assumption that $cov(\hat{\mu}_{st}, \nu_{ist}) = 0$. Transitory VA ($\hat{\eta}$) is by construction orthogonal to persistent VA ($\hat{\mu}$), thus ignoring $\hat{\eta}$ will not cause an omitted-variable bias in the estimate of γ_1.

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 can 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).

Table 1: Standard deviations of in-compulsory school outcomes and VA indicators

	Written exam score	Oral exam score	Teacher grades
Student-level SD	1	1	1
SD of school*year-mean	0.298	0.258	0.261
SD of VA indicator	0.093	0.099	0.127
R^2 from regression on X	0.199	0.155	0.309

Note: Table shows student-level standard deviations for each outcome in the first row, student-weighted standard deviations of school means in the next two rows and R^2 from a student-level regression of the outcome on the background variables in the last row. 2004-2018 graduation cohorts.

4 Peristent school VA and long-term student 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 indicators are also predictive of longer-term outcomes.

VA is estimated from students graduating from 2004-2018, while the 2002 and 2003 cohorts are reserved for testing the indicators. This allows me to test how the indicators predict outcomes around the age of 31-32. I construct VA indicators for three outcomes: end-of-compulsory written exam scores, oral exam scores and teacher grades. In Table 1 I show the dispersion of the different outcomes and indicators. Figure A6 in the Appendix shows the distribution of the school-by-year means and VA for written exam scores. All outcomes are standardized within cohort at the student level. The school-by-year averages have (student-weighted) standard deviations ranging from 0.25 to 0.30 student-level standard deviations, higher for written-exam score than the other outcomes. The VA indicators have standard deviations around 0.09-0.13, higher for teacher grades than the exam scores. The reduced dispersion of the VA indicators compared to the school-by-year averages reflect both averaging over cohorts and adjustment for between-school-and-cohort differences in student composition. However, the student background variables only have moderate explanatory power at the individual level, ranging from 16 percent for oral exam scores to 31 percent for teacher grades.

The main questions are whether the value-added indicators are able to forecast measures of in-school performance and whether the indicators capture variation in competence that are restricted to exam scores, or to what extent school value-added also predict students' later outcomes. Table 2 shows associations between estimated VA and different outcomes. Each cell reports the key coefficient from a separate regression corresponding to (4), 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 written 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 written exam scores out of sample. I find a coefficient close to one, and although slightly attenuated, not significantly different from one. 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 written exam VA indicator and oral exam scores or teacher grades is similar to the the corresponding individual-level relationships (shown in the fourth row).

There is consistently a highly significant and strong association between the written exam-based indicator and later outcomes. The next two measures, on-time completion of the first year of upper secondary and graduating from upper secondary within five years, are both more strongly associated with the exam-based indicator than with own exam score.

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. Nine percent of all students are still under education 16-17 years after completion of compulsory schooling, and students with higher exam scores more often. However, the VA indicator is negatively related to whether the student still is in education. The VA indicator is also related to labor market outcomes. The share employed[11] 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 associations with exam-based VA indicators are more than twice as strong as the associations with individual-level exam scores.

A school-level one-standard deviation difference in exam value-added (i.e. a .093 student-level SD difference), corresponds to a predicted difference of 2.9 percentage points in upper secondary completion, a 0.5 percentage points difference in labor

[11]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.

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 (NOK 100k)	(11) Log earnings
$\hat{\mu}$ written exam	0.919**	0.655**	0.762**	0.208**	0.309**	0.813**	-0.026**	0.077**	-0.058**	1.599**	0.199**
	(0.057)	(0.058)	(0.066)	(0.023)	(0.024)	(0.159)	(0.012)	(0.015)	(0.014)	(0.155)	(0.023)
$\hat{\mu}$ oral exam	0.677**	0.969**	1.026**	0.091**	0.143**	0.610**	-0.004	0.041**	-0.037**	1.267**	0.175**
	(0.061)	(0.056)	(0.062)	(0.024)	(0.026)	(0.197)	(0.011)	(0.014)	(0.012)	(0.161)	(0.024)
$\hat{\mu}$ teacher	0.437**	0.589**	1.007**	0.100**	0.151**	0.508**	-0.006	0.052**	-0.047**	1.097**	0.129**
	(0.045)	(0.044)	(0.042)	(0.018)	(0.019)	(0.150)	(0.009)	(0.010)	(0.009)	(0.118)	(0.017)
Written exam score	1.000	0.514**	0.643**	0.123**	0.153**	0.944**	0.013**	0.035**	-0.035**	0.644**	0.087**
	(.)	(0.004)	(0.004)	(0.002)	(0.002)	(0.015)	(0.001)	(0.001)	(0.001)	(0.014)	(0.002)
Oral exam score	0.492**	1.000	0.606**	0.113**	0.150**	0.881**	0.012**	0.037**	-0.035**	0.631**	0.083**
	(0.004)	(.)	(0.003)	(0.002)	(0.002)	(0.015)	(0.001)	(0.001)	(0.001)	(0.012)	(0.002)
Teacher grade	0.708**	0.698**	1.000	0.183**	0.236**	1.309**	0.013**	0.059**	-0.056**	0.866**	0.110**
	(0.004)	(0.003)	(.)	(0.002)	(0.002)	(0.015)	(0.001)	(0.002)	(0.002)	(0.014)	(0.002)
N	83372	83372	83145	82723	82723	81154	83372	83372	83372	81135	72504
# clusters	2026	2026	2025	2026	2026	2025	2026	2026	2026	2025	2012
$\bar{y}$	0.015	0.015	0.037	0.808	0.714	13.909	0.089	0.870	0.114	5.069	1.642

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 2019, i.e. 16-17 years after graduation from compulsory school, around age 31-32. (6) is nominal duration of highest completed degree (in years, including compulsory school); (7) is a dummy for whether the person in education in 2019; (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 2004-2018 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-by-year level. Significant at * 10%, ** 5%

market participation and a 1.9 percent earnings difference (given positive earnings). The 90-10 percentile difference in VA is 0.23 student-level SDs, corresponding to a 4.6 percent earnings difference. 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.

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

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 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 written exam-VA indicators, although not all differences are statistically significant. The oral exam-indicators are mostly in-between. As shown in Table 1, the standard deviation of the teacher grade VA-indicator is greater than that of the exam score indicators. However, for all post-school outcomes in Table 2 except NEET, the difference in outcomes associated with a one-SD difference in the exam score indicator is greater than the difference in a one-SD difference in the teacher grade indicator.

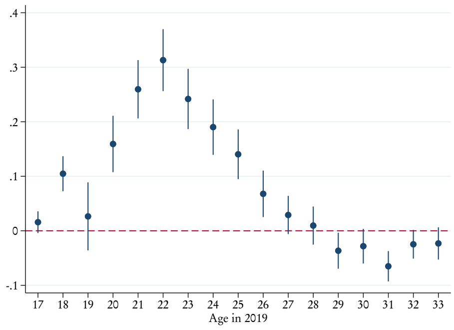
Thus, indicators constructed from oral exam scores or teacher grades also capture persistent differences between schools. These indicators are also predictive of later outcomes, but less so than indicators based on written exam scores, in particular teacher grade VA. This is despite the fact that student-level 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. The strong associations between exam score VA and longer-term outcomes compared to the student-level associations of outcomes and exam scores or teacher grades suggests that a given contribution of a school can more than make up for a similar-sized disadvantage in terms of student background (keeping in mind that the dispersion of schools contributions is of course much smaller than the dispersion in student background, cf. Table 1).

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 weaker association between teacher-grade VA and later outcomes suggests that this mechanism is not very important.

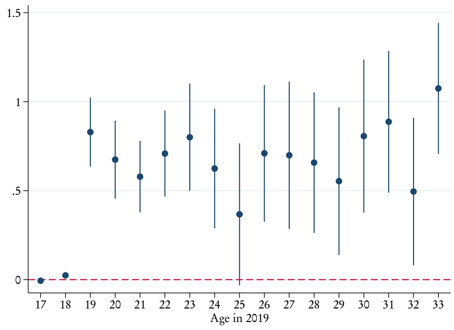
High- and low-VA schools may be located in different communities, which may differ also in other way, e.g. in terms of local labor markets. Thus, differences in later outcomes may not reflect differences in VA. In Table A1 in the Appendix I reproduce Table 2, but with municipality fixed effects. The associations are mostly similar or stronger than 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.

In Table 2 I restricted the sample to the 2003 and 2004 cohorts, to avoid overlap with the cohorts used to construct the indicators. In Figure 1 I remove this restriction, in order to see how the associations between exam score VA and long-term outcomes vary with age. All the long-term outcomes are observed in 2019, thus students aged 32 and 33 are the 2003 and 2004 cohorts studied in Table 2, while younger ages correspond to later cohorts. While these cohorts have contributed to the estimation of the VA indicators, they still do not contribute directly to the indicator for their own graduation year, cf. Section 3.

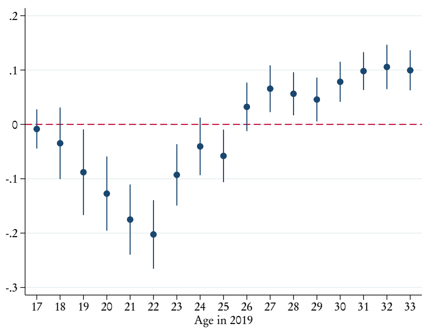
Panel (a) of Figure 1 shows the association between VA and being in education. Almost all Norwegian students start in upper secondary after completing compulsory, thus it is unsurprising that there is no effect on being in education at age 17. However, at age 18, corresponding to the second year of upper secondary, there is already a difference in educational participation between students from high- and low-VA schools. The association peaks during the early 20's, and is reversed around age 30, possibly reflecting some later catching up of the students from low-VA schools. However, as can be seen from panel (b), there is no evidence for catching up in terms of completed years of schooling. Panel (c) shows labor market participation. This largely mirrors educational participation, with fewer students from high-VA schools working at ages where more are in education. However, students from high-VA schools have a persistently higher labor market participation from the



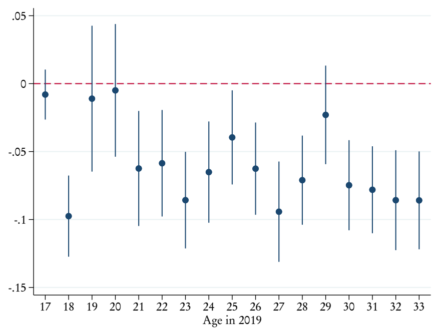
(a) In education



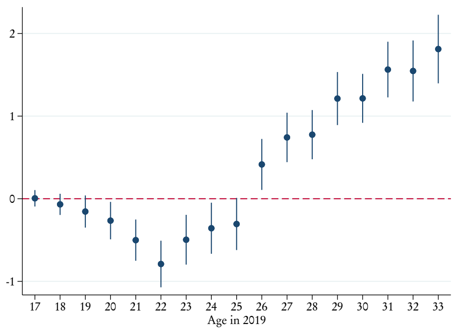
(b) Years of schooling



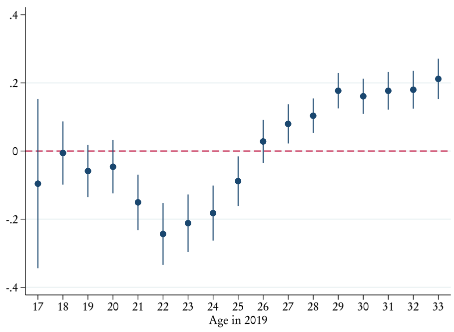
(c) Earnings > G



(d) Inactive



(e) Earnings



(f) Log earnings

Figure 1: Associations between exam score VA and outcomes by age at observation
Note: Estimated associations between exam score VA and outcomes by age in 2019 with confidence intervals. VA constructed from 2004-2018 cohorts. Standard errors adjusted for school-by-cohort clustering.

late 20's onward. Also, the associations of VA with educations and employment do not fully cancel out, as students from high-VA schools have persistently lower levels of inactivity from around age 20.

Finally, panels (e) and (f) show associations with absolute and relative labor earnings. Earnings will reflect both labor market participation and wages, which in turn will reflect skills and qualifications. Panel (e) clearly shows that while the higher participation in education in the early 20's contributes to lower earnings for students from high-VA schools, this loss is dominated by the earnings gain from the late 20's onward. While panel (e) shows diverging earnings, panel (f) suggests that this largely reflects increasing earning levels over the life cycle, as relative earnings gains from graduating from a high-VA school stabilize from around age 30.

5 Different VA estimates: Persistent and transitory VA and different set of controls

Studying longer-term educational and labor market outcomes as in the previous section I am constrained to use cohorts before the introduction of 5th and 8th grade test score. In the current section I will use more recent cohorts, for whom standardized test in grades five and eight are available, to study indicators using different sets of control variables. I will also investigate transitory VA (cf. discussion in Section 3), and compare its dispersion with that of persistent VA. Like in the previous section I will estimate VA indicators for a set of cohorts, in this case the 2015-2019 graduation cohorts, and reserve earlier cohorts for testing the out-of-sample performance of the indicators.[12]

In the following, I study three different outcomes: end-of-compulsory written exam score, grade eight test score and grade five test score. For the first two outcomes I will construct VA indicators controlling either only for family background or also for the previous test scores: grade eight tests for the exam score indicators and grade five test when studying grade eight tests. As noted in Section 3, different controls change the interpretation of the indicators. For grade five tests there are no prior tests, thus I will only control for family background. Finally, for each VA indicator I will estimate persistent VA (μ, cf. Section 3) and transitory VA (η).

In Table 3 I present SDs of test scores, school-by-year means and the different VA indicators, similar to what I did for the VA indicators constructed from the 2005-2018 cohorts in Table 1. In Figures A7-A9 in the Appendix I show the corresponding distributions. All outcomes are standardized at the student level, and thus

[12]The 2013 cohort is the first for whom grade five tests are available, while the 2015 cohort is the last for whom completion of upper secondary can be observed. Thus, for the 2013-2014 cohorts I can relate outcomes including completion of upper secondary to own controls including grade five and eight test scores and VA indicators estimated from later cohorts.

Table 3: Standard deviations of outcomes and VA indicators throughout compulsory school

	Written exam score	Grade 8 test	Grade 5 test
Student-level	1	1	1
School-by-year mean	0.296	0.302	0.308
Indicators controlling only for family background			
Persistent VA (μ)	0.083	0.114	0.134
Transitory VA (η)	0.195	0.202	0.245
Indicators controlling also for pre-test			
Persistent VA (μ)	0.072	0.074	
Transitory VA (η)	0.170	0.150	

Note: Table shows student-level standard deviations for each outcome in the first row, student-weighted standard deviations of school means in the next rows. 2015-2019 graduation cohorts.

have comparable scales. Furthermore, all outcomes have school-by-year means close to 0.3, similar to those of written exam scores in Table 1. Restricting the sample to more recent cohort reduces the standard deviation of persistent exam score VA somewhat, from 0.093 to 0.083. The dispersion of VA indicators estimated from the grade eight and five test scores are greater, with standard deviations of 0.0114 and 0.134 student-level standard deviations. There are approximately twice as many primary schools with grade five VA indicators as lower secondary schools with exam score indicators, likely contributing to greater school-level dispersion. However, the grade eight scores are associated with lower secondary schools, even if the outcome is essential end-of-primary proficiency. Thus, the aggregation of the grade eight and exam score indicators is the same, indicating that there is greater dispersion of VA among primary schools, even when the schools are aggregated to the students' lower secondary schools.[13] The lower part of Table 3 shows results for indicators controlling for previous test scores. VA for exam scores controlling for grade eight test scores has slightly lower dispersion than when controlling only for family background. The dispersion of VA for grade eight tests when controlling for the previous test is very similar the dispersion of exam score VA.

Table 3 also shows the standard deviation of transitory VA, i.e. school-by-year means of student performance not explained neither by student characteristics nor persistent VA. These standard deviations are consistently two to three times larger than the standard deviations of persistent VA, implying that persistent VA explains 11-25 percent of the unexplained variation in school-by-year mean outcomes. Transitory VA may reflect some combination of school characteristics that vary between

[13] Alternatively, the higher dispersion may reflect a poorer ability of the of controls to account for differences between students.

Table 4: Predicting exam scores of 2013-2014 graduates

	(1)	(2)	(3)	(4)	(5)	(6)
	Indicators control for previous score			Indicators control for background only		
Exam value-added	1.010**	0.966**		1.010**		
	(0.085)	(0.072)		(0.059)		
8th grade value-added	0.646**		0.635**		0.485**	
	(0.065)		(0.068)		(0.045)	
5th grade value-added	0.238**					0.239**
	(0.031)					(0.035)
Controls (cubic + school mean):						
Grade 8 score		Yes				
Grade 5 score			Yes			
Family background	Yes	Yes	Yes	Yes	Yes	Yes
N students	104393	104393	104393	104393	104393	104393
N clusters	1076	1076	1076	1076	1076	1076
R^2	0.203	0.463	0.401	0.203	0.199	0.197

Note: Outcome is exam score of 2013-2014 cohorts. Value-added indicators are constructed from the 2015-2019 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%

successive cohorts and average unobserved characteristics of the students. In the Norwegian context, different teachers teaching different cohorts is likely to contribute to the former. In the next section I will investigate quasi-experiments that allow me to distinguish better between alternative explanations for the differences in transitory VA.

In Table 4 I further investigate the out-of-sample predictive power of the different indicators, similar to what I did in Table 2. I regress student-level exam scores of students completing compulsory school in 2013-2014 on VA indicators constructed from the 2015-2019 cohorts, controlling for the same control variables used in the construction of the indicators but for the 2013-2014 students themselves.

As highlighted by equations (3) and (4), 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 4 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 a specification that distinguishes between VA before grade 5, from grade 5 to 8 and from 8 to the end of compulsory school, as well as students' (pre-school) background.[14] On average,

[14] Students' background is partly decided at birth (sex and immigrant background) and partly

differences in from 8th grade to exam scores predict exam scores without bias. The estimated coefficient on the value-added indicator is precisely estimated and not significantly different from one. Value-added in from 5th to 8th and before 5th grade 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 Table 3 we also know that the dispersion 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 0.65 SD and 0.24 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, but 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.

For the 2013-2014 graduates I have data both on test scores and on completion of upper secondary school. In Table 5 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 4. 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 later in compulsory school is more strongly related to completion than VA from earlier stages, as for exam scores in Table 4, However, the gradient is less pronounced than in Table 4.

measured at age 16 (parents' education). However, even if parents' formal education may change during the students' childhood, this is relatively rare. Furthermore, in the relationship with their children's school performance, parents' education is also a proxy for fixed characteristics of the parents thatcorrelated with education. While the background variables themselves reflect pre-school characteristics, the relationships with exam scores may change over time. As highlighted by eqs. (3) and (4) the coefficients on student background will represent the total association between background and school performance.

Table 5: Predicting completion of upper secondary for 2013-2014 graduates

	(1)	(2)	(3)	(4)	(5)	(6)
	Indicators control for previous score			Indicators control for background only		
Exam value-added	0.137**	0.120**		0.176**		
	(0.032)	(0.029)		(0.023)		
8th grade value-added	0.108**		0.100**		0.109**	
	(0.023)		(0.022)		(0.015)	
5th grade value-added	0.065**					0.065**
	(0.011)					(0.011)
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	107252	107252	107252	107252	107252	107252
N clusters	1080	1080	1080	1080	1080	1080
R^2	0.056	0.122	0.094	0.056	0.055	0.055

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%

In Figure 2 I present results from regressing exam scores on value-added, similar to columns (2) and (4) of Table 4, for each of the years 2010--2019. As in Table 4 VA are based on the 2015-2019 cohorts. Unsurprisingly, for each year of the estimation period the coefficients are close to one. Before the estimation period The coefficients are close to one in the two years immediately preceding the estimation period, but are smaller than one at greater time differences. . However, three to five years before the estimation period, the coefficients are still 0.7-0.9, and the confidence intervals overlap with the coefficients of the previous years. Thus, the VA capture characteristics of schools that have substantial persistence, in addition to having persistent impacts on students.

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

From the previous sections 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.

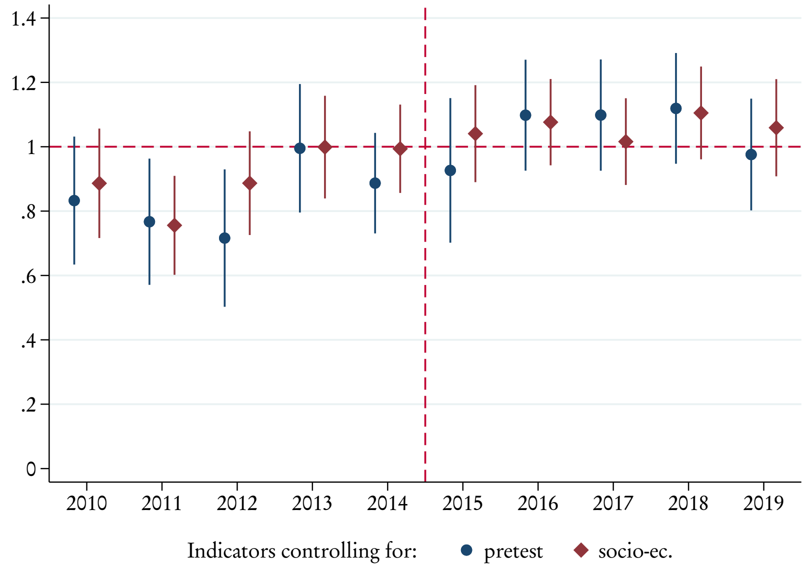


Figure 2: Exam score and estimated school quality by cohort
Note: The graph show the estimated relationship between exam score and VA indicators, similar to columns (2) and (4) of Table 4. VA indicators are based on the 2015-2019 cohorts.

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. In each of these situations I will estimate VA based on students/neighborhoods not changing schools (stayers) and see how outcomes of the students/neighborhoods that do change schools (movers) are related to these estimates. Is likely not random which students or neighborhoods change schools. However, I will discuss whether VA of the new school can be considered conditionally good-as-random, and thus whether the we can estimate effects of the new (and potentially) the old school on the movers.

6.1 School changers

While Norway lack a central registry of which school compulsory school students attend, 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. (6)), conditional on the old school, this will give a consistent estimate of the effect of the value-added of the new school.

In Table 6 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 based on indicators controlling for lagged achievement, such that VA is VA for lower secondary, conditional on achievement at the start of lower secondary, as in eq. (3). In the lower panel I use indicators controlling only for family background, such that VA is the combined VA of primary and lower secondary, as in eq. (4). 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. In the Norwegian context, where successive cohorts will often be taught by different teachers, this likely reflects, at least in part, within-school differences in teacher VA, cf. Chetty et al. [2014b]. 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 6. 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. The significant effect of transitory VA on exam scores in column (1) indicate that transitory VA affect students' skills. The absence of effect in column (2) may reflect relative grading in the new school, where teachers do not recognize that the cohort is high-performing relative to previous cohorts. Eventually, exam scores provide schools with a signal of their students relative performance, which may allow teachers to adjust their grading to persistent differences. However, teachers do not yet have their students' exam score at the time of setting teacher grades and thus may be unaware of year-to-year variation. 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 6 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

Table 6: Exam score 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						
$\hat{\mu}^{Old}$	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 2011-2019). Indicators are constructed from students in the same cohorts that don't change schools. Outcomes in columns (1)-(5) are the same as the corresponding outcomes 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%.

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. Thus, as there is no indication of systematic sorting of students, I conclude that the results in Table 6 provide credible estimates of the effects of the quality of the new school, as measured by VA, on the outcomes of students changing schools.

In the lower panel of Table 6 I show similar results using indicators controlling only for student background, while I control for student background but not lagged test scores 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 A2 in the Appendix I present results similar to those in Table 6, but where I control flexibly for the grade 8 school with school dummies. The results for the new school VA are very similar to those in Table 6.

In Table A3 in the Appendix I repeat the analyses in Table 6 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.

6.2 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.[15] 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

[15] 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).

upper secondary school. While some students attend the same school throughout compulsory school, many schools are only primary or upper secondary schools, and 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 7 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.[16] Thus, if VA of the school actually attended by the movers is uncorrelated with the VA of their predicted school when these schools are not the same, 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

[16]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 graduate from, 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.

school are significantly related to teacher grades, completion of grade 11 and high school and NEET status and earnings 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 6, transitory VA of the new neighborhood is much more strongly related to exam scores than to teacher grades, suggesting relative grading. 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 of the new neighborhood 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).

In Table A4 in the Appendix I present results similar to Table 7, 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

Table 7: 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 in columns (1)-(7) are the same as in Table 6. $\hat{\mu}^{Old}$ and $\hat{\eta}^{Old}$ are persistent and transitory VA (exam scores adjusted for student background) of the modal lower secondary school of the student's neighborhood at school-starting age and $\hat{\mu}^{New}$ and $\hat{\eta}^{New}$ are similarly the VA of the modal 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%

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.

6.3 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 sorting and the context suggests that sorting based on value-added is unlikely, students moving may do so in a way that creates a correlation between value-added and unobserved characteristics of the 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 A10 and A11 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.

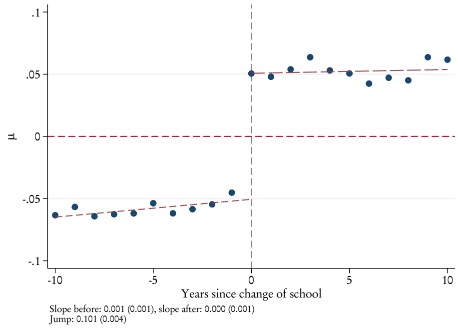
A challenge interpreting the results from these analyses is that I don't observe the process leading up to and following the school change, I only observe that students from a given neighborhood complete one school before and another after a given year. This can reflect rezoning of existing schools (some 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 graduating a few years after their neighborhood changed school 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 discontinuous 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.

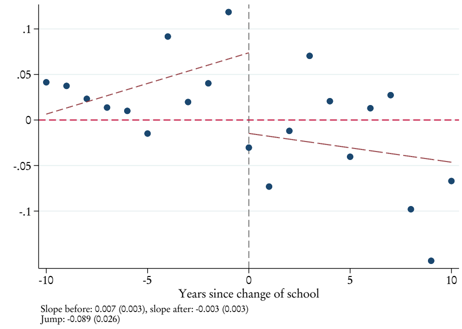
In sub-figure (c) I show a similar event study using average exam scores. Average exam scores change by .014 SD, in the same direction as the change in predicted VA, but the change is not significant. Finally, sub-figure (c) shows the event study for residualized exam scores, constructed by adjusting 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 8 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 .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.

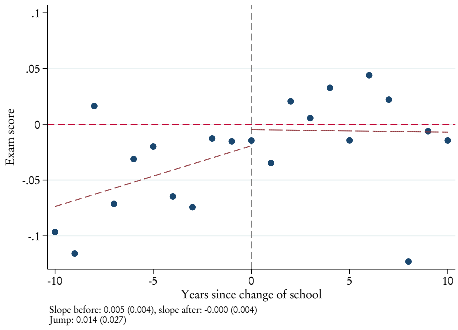
The results for teacher grades, in column (2), mostly reflect those for exam scores, although the coefficients are smaller. The most striking difference is that teacher grades are unrelated to transitory VA for graduates from the new school, and only weakly related to transitory VA for graduates from the old school, likely



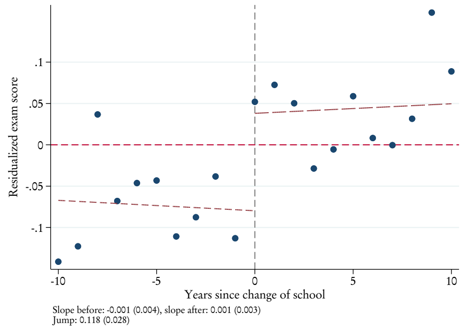
(a) Predicted persistent VA



(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 VA and average outcomes 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

Table 8: Effect of change in 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%

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.

In Table A5 in the Appendix I repeat the analyses in Table 8 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.

7 Conclusion

Schools are a key instrument of policy makers to foster skills and provide all children with opportunities. It thus of great relevance to identify schools that do this to a greater or smaller extent. In this paper, I study school quality in Norwegian compulsory school. Previous studies have found important differences in school VA in the US, I find persistent differences in VA also in Norway, with important consequences for their students' long-term outcomes.

Estimating persistent school quality as shrinkage-adjusted VA estimates adjusting exam scores for family background, I find that these estimates are forecast-unbiased for in-school outcomes and strongly associated with longer-term outcomes, including outcomes in the labor market. Three quasi-experiments, where students move/change school or the link between neighborhood and school is changed, allow 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 violated. I thus conclude that the persistent VA measures are good measures of school quality.

Compared to the VA indicators based on exam grades, indicators based on teacher grades are much less informative about outcomes other than teacher grades. This shows 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. As GPA mostly is based on teacher grades, this is also evidence that high-VA lower secondary schools impact long-term outcomes mostly by providing skills, not by giving their students an advantage when applying for upper secondary schools.

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 scope for 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.

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 Pathak, and Christopher Walters. Interpreting tests of school vam validity. *American Economic Review*, 106(5):388–92, 2016.

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.

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. Using school choice lotteries to test measures of school effectiveness. *American Economic Review*, 104(5):406–11, 2014.

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.

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.

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

Eric A Hanushek and Steven G Rivkin. Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2):267–71, 2010.

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.

OECD. *Education at a Glance 2006*. 2006. doi: https://doi.org/https://doi.org/10.1787/eag-2006-en. URL https://www.oecd-ilibrary.org/content/publication/eag-2006-en.

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.

Jesse Rothstein. Measuring the impacts of teachers: comment. *American Economic Review*, 107(6):1656–84, 2017.

Appendix

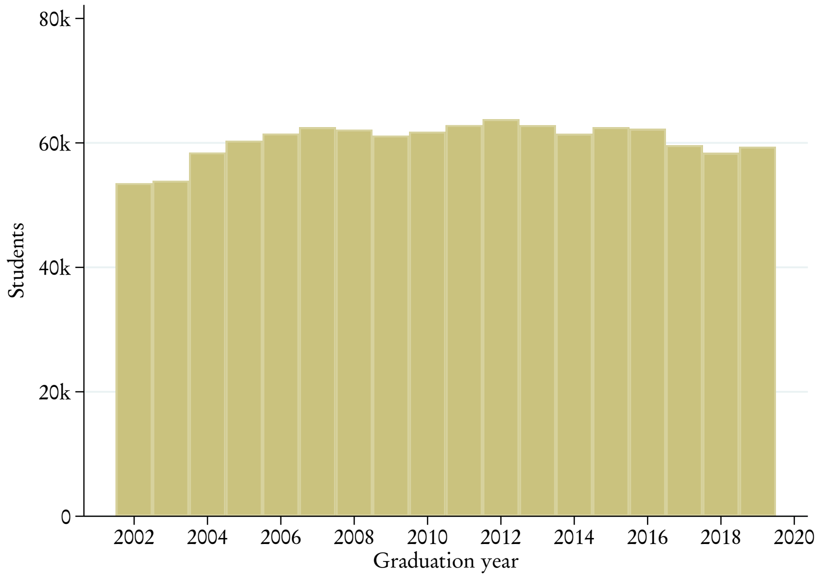


Figure A1: Cohort sizes

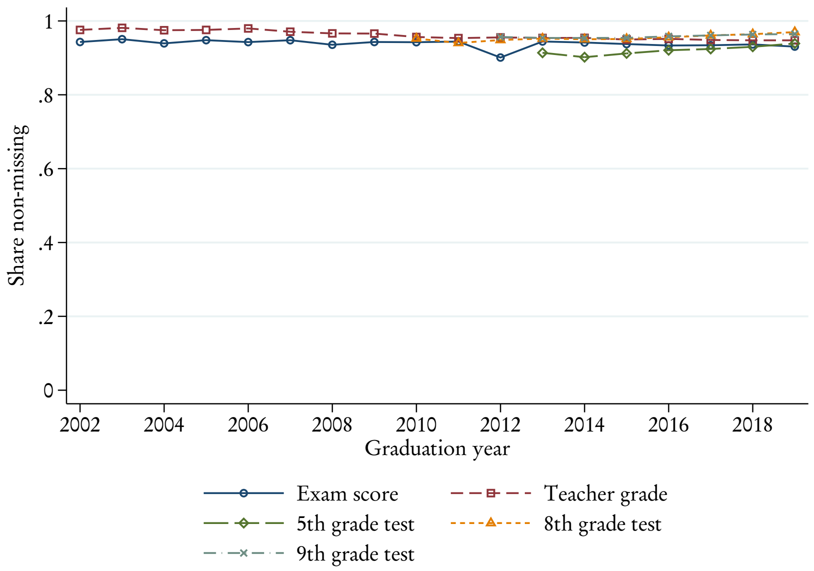


Figure A2: Share of students with non-missing exam scores, teacher grades and test scores by cohort

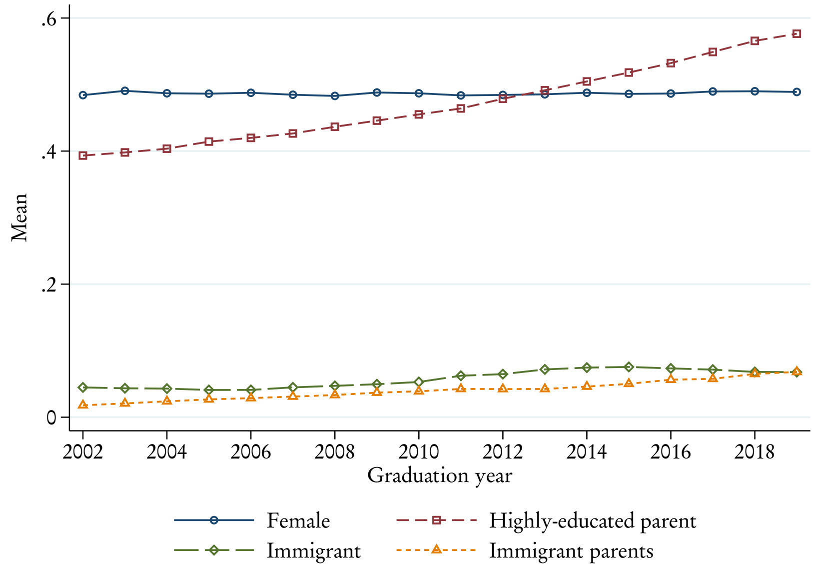


Figure A3: Mean background characteristics by cohort

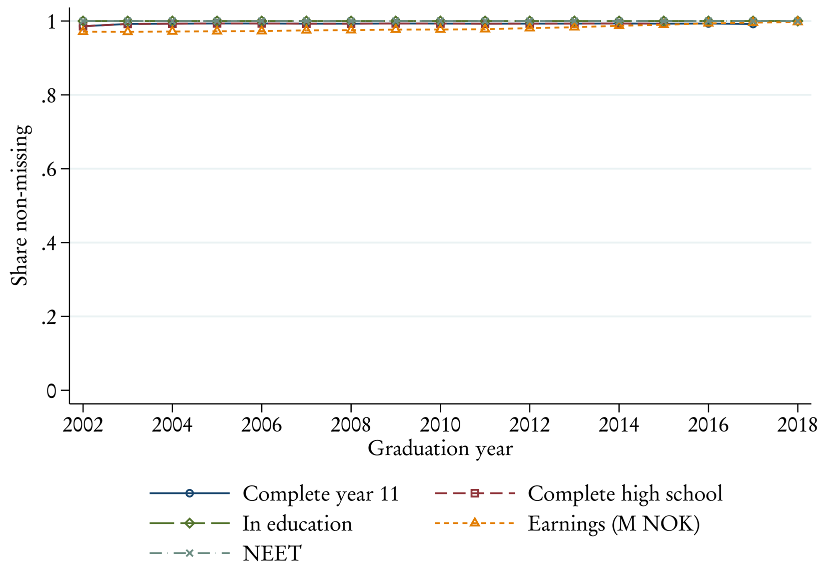


Figure A4: Share of students with non-missing longer-term outcomes by cohort

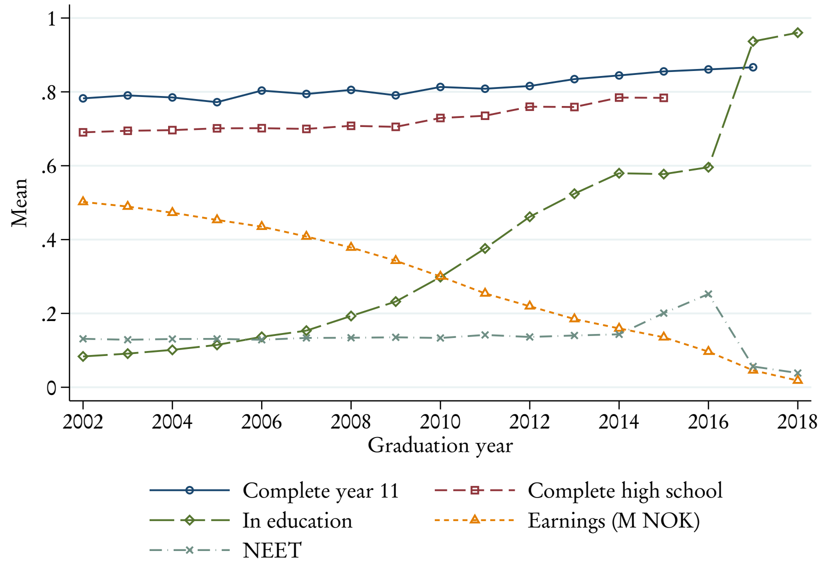


Figure A5: Mean longer-term outcomes by cohort

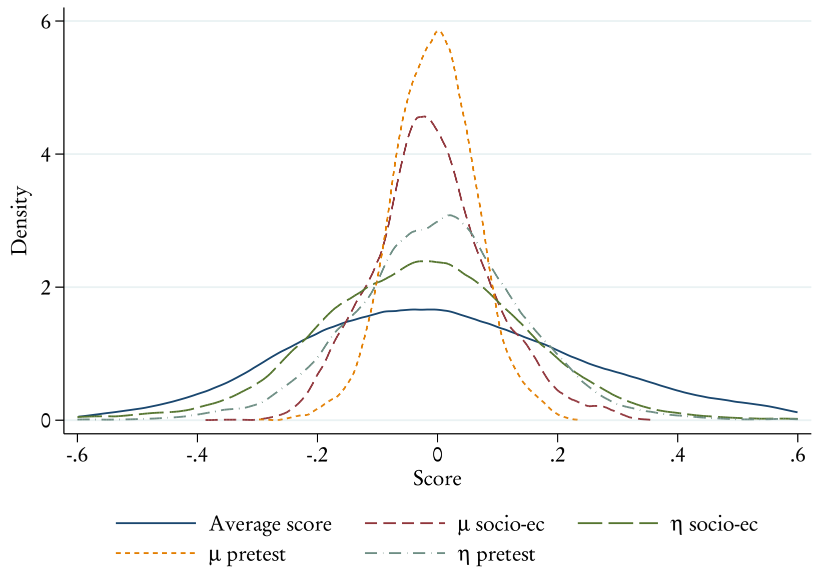


Figure A6: School-by-year average exam score and estimated VA for the 2004-2018 cohorts

Table A1: 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 (NOK 100k)	(11) Log earnings
$\hat{\mu}$ written exam	0.990**	0.903**	0.904**	0.291**	0.283**	1.263**	-0.002	0.096**	-0.073**	1.586**	0.214**
	(0.094)	(0.093)	(0.134)	(0.062)	(0.053)	(0.327)	(0.019)	(0.030)	(0.026)	(0.336)	(0.036)
$\hat{\mu}$ oral exam	0.920**	1.169**	1.254**	0.284**	0.328**	1.188**	-0.006	0.073**	-0.068**	1.660**	0.208**
	(0.141)	(0.137)	(0.173)	(0.093)	(0.087)	(0.253)	(0.014)	(0.030)	(0.029)	(0.567)	(0.061)
$\hat{\mu}$ teacher	0.555**	0.697**	1.095**	0.188**	0.203**	0.888**	0.004	0.065**	-0.064**	1.011**	0.109**
	(0.109)	(0.122)	(0.088)	(0.054)	(0.065)	(0.235)	(0.011)	(0.022)	(0.021)	(0.367)	(0.039)
Written exam score	1.000	0.514**	0.645**	0.123**	0.151**	0.938**	0.012**	0.035**	-0.035**	0.643**	0.087**
	(.)	(0.004)	(0.005)	(0.002)	(0.002)	(0.018)	(0.001)	(0.002)	(0.002)	(0.018)	(0.003)
Oral exam score	0.491**	1.000	0.605**	0.115**	0.150**	0.881**	0.012**	0.037**	-0.036**	0.631**	0.083**
	(0.004)	(.)	(0.005)	(0.002)	(0.002)	(0.018)	(0.001)	(0.002)	(0.002)	(0.015)	(0.003)
Teacher grade	0.716**	0.701**	1.000	0.186**	0.238**	1.327**	0.013**	0.060**	-0.057**	0.868**	0.110**
	(0.005)	(0.003)	(.)	(0.003)	(0.002)	(0.018)	(0.001)	(0.002)	(0.002)	(0.028)	(0.004)
N	83372	83372	83145	82723	82723	81154	83372	83372	83372	81135	72504
# clusters	428	428	428	428	428	428	428	428	428	428	428
$\bar{y}$	0.015	0.015	0.037	0.808	0.714	13.909	0.089	0.870	0.114	5.069	1.642

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%

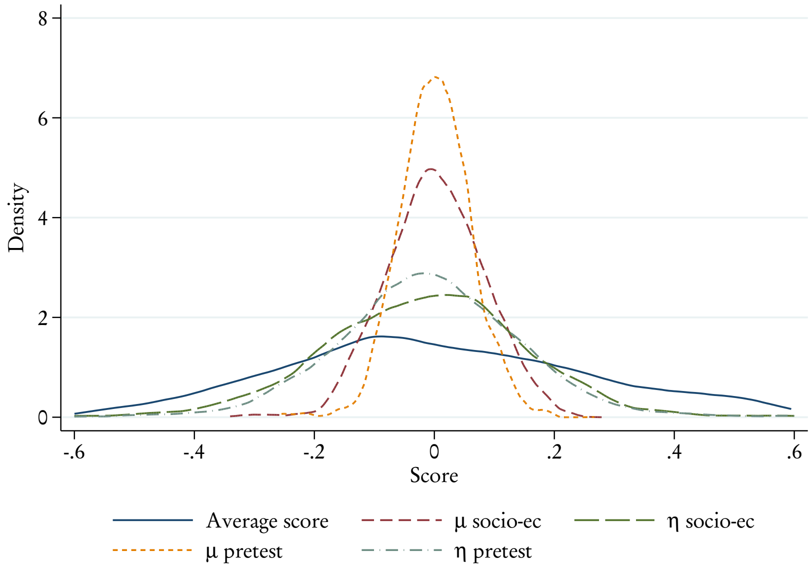


Figure A7: School-by-year average exam score and estimated VA for the 2015-2019 cohorts.

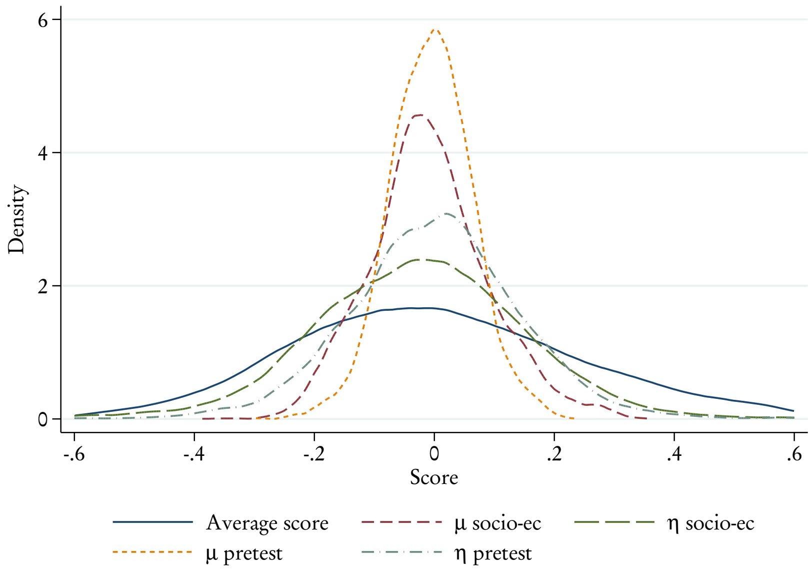


Figure A8: School-by-year average grade eight test score and estimated VA for the 2015-2019 cohorts

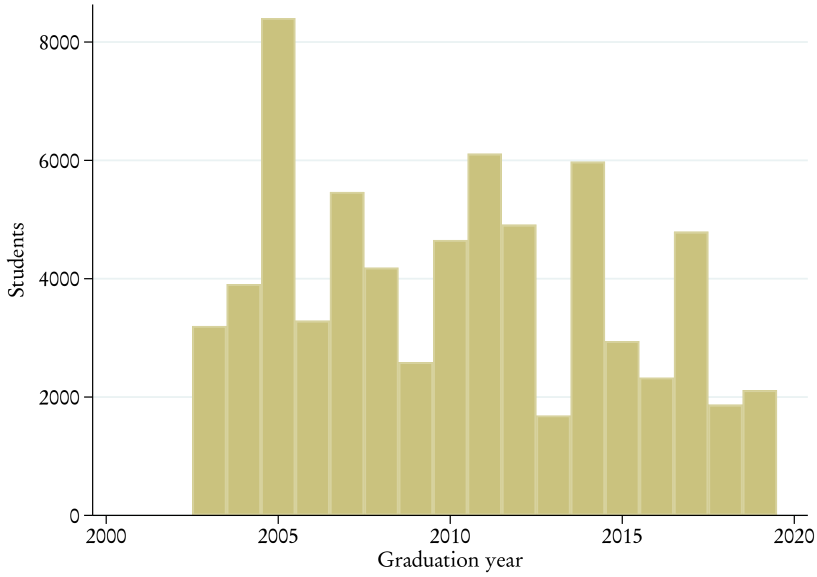


Figure A10: Year of change predicted school

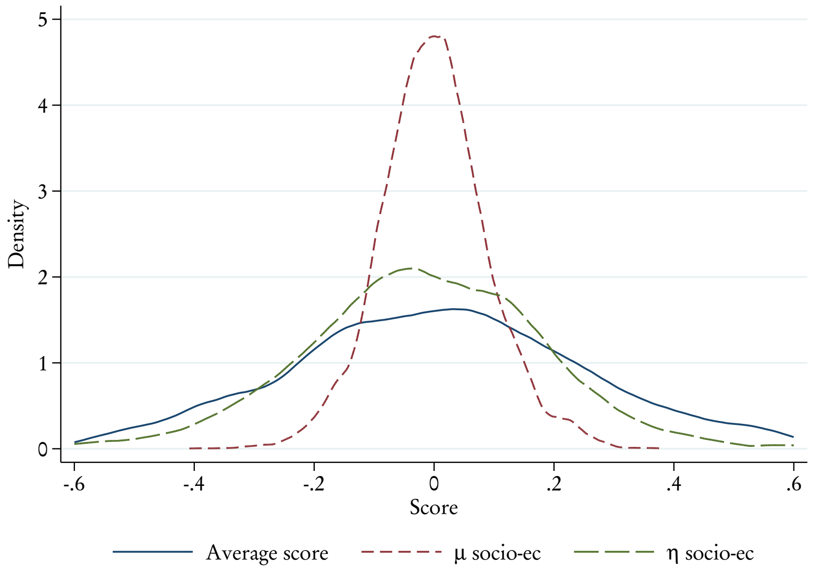


Figure A9: School-by-year average grade five test score and estimated VA for the 2015-2019 cohorts

Table A2: 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	-0.749	-0.628	0.281	1.768	-1.424	2.511
	(2.290)	(2.723)	(1.465)	(1.955)	(1.314)	(2.143)
$\hat{\mu}$ year 9-school	0.818**	0.460**	0.068	0.192	0.022	-0.197
	(0.142)	(0.151)	(0.077)	(0.121)	(0.070)	(0.153)
$\hat{\eta}$ year 8-school	0.065	0.016	0.009	0.199	-0.135	0.242
	(0.192)	(0.229)	(0.124)	(0.163)	(0.108)	(0.174)
$\hat{\eta}$ year 9-school	0.291**	-0.017	-0.024	-0.012	-0.020	-0.008
	(0.052)	(0.047)	(0.025)	(0.041)	(0.022)	(0.052)
Indicators controlling for family background						
$\hat{\mu}$ year 8-school	1.164	0.688	0.769	2.210	0.233	1.155
	(1.725)	(1.988)	(0.962)	(1.484)	(0.954)	(1.177)
$\hat{\mu}$ year 9-school	0.724**	0.463**	0.025	0.115	-0.018	0.001
	(0.105)	(0.115)	(0.057)	(0.081)	(0.045)	(0.066)
$\hat{\eta}$ year 8-school	0.223	0.123	0.063	0.265*	-0.010	0.098
	(0.171)	(0.196)	(0.094)	(0.147)	(0.090)	(0.113)
$\hat{\eta}$ year 9-school	0.302**	0.074	0.007	0.017	-0.016	0.003
	(0.050)	(0.052)	(0.024)	(0.039)	(0.021)	(0.027)
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 6. Significant at * 10%, ** 5%

Table A3: 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	0.018	0.062	-0.018	0.029	0.011	0.053
	(0.088)	(0.081)	(0.045)	(0.070)	(0.040)	(0.085)
$\hat{\mu}$ year 9-school	-0.029	0.934**	-0.069	0.059	0.013	0.001
	(0.084)	(0.088)	(0.044)	(0.069)	(0.041)	(0.084)
$\hat{\eta}$ year 8-school	0.132**	0.159**	0.045	0.053	-0.017	-0.028
	(0.053)	(0.054)	(0.028)	(0.040)	(0.023)	(0.048)
$\hat{\eta}$ year 9-school	-0.042	0.471**	-0.002	0.043	-0.001	0.094*
	(0.051)	(0.051)	(0.028)	(0.041)	(0.026)	(0.051)
Indicators controlling for family background						
$\hat{\mu}$ year 8-school	0.151*	0.189**	0.022	0.079	0.021	0.062
	(0.083)	(0.085)	(0.040)	(0.063)	(0.036)	(0.046)
$\hat{\mu}$ year 9-school	0.297**	0.980**	-0.008	0.114*	-0.028	0.028
	(0.081)	(0.084)	(0.040)	(0.062)	(0.035)	(0.043)
$\hat{\eta}$ year 8-school	0.061	0.118**	0.044*	0.051	-0.028	0.034
	(0.049)	(0.053)	(0.025)	(0.036)	(0.020)	(0.028)
$\hat{\eta}$ year 9-school	0.046	0.474**	0.041*	0.078**	-0.018	0.018
	(0.048)	(0.050)	(0.024)	(0.037)	(0.023)	(0.025)
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 6. Significant at * 10%, ** 5%

Table A4: 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 6.$\hat{\mu}^{Old}$ 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%

Table A5: 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%

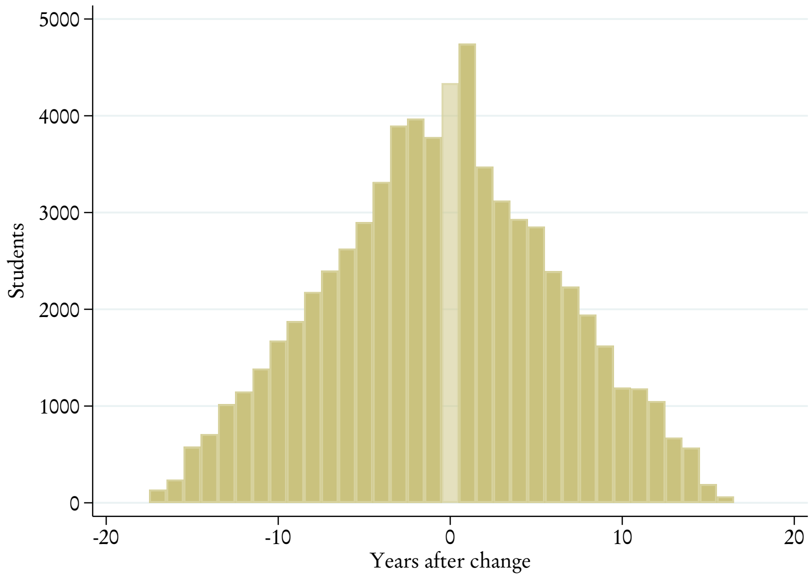


Figure A11: Years since change predicted school

Table A6: 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%