CESifo WORKING PAPERS

9769
2022

May 2022

School Value-Added and Long-Term Student Outcomes

Lars J. Kirkebøen

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9769*

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 the value-added (VA) of Norwegian schools, 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 corresponds to 1.9 percent higher earnings at around age 32. Three quasi-experiments using variation from student mobility and changes in neighborhoods' assignment to schools indicate that the differences captured by the VA indicators do indeed reflect differences in school quality, rather than unobserved student characteristics. Analyses of teacher grades and exam scores suggest that the former are heavily influenced by relative grading, and that the effect of exam score VA on long-term outcomes reflects the effects of competencies and skills acquired in school. In addition to shedding light 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.

JEL-Codes: Q540, D700, D500.

Keywords: climate economics, international environmental agreements, coalition formation, heterogeneous countries, integrated assessment models.

Lars J. Kirkebøen
Statistics Norway
Research Department
Oslo / Norway
kir@ssb.no

May 2022
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.

1 Introduction

Primary and secondary schooling has 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. 2014a). This has further spurred interest in interventions that improve student performance (Fryer, 2017) and in identifying effective teachers and schools. There is substantial recent research showing that the contributions to students' learning varies across schools, and that these contributions can be predicted accurately using value-added (VA) models. There is active research into how available data can best 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] Recent years have seen increasing interest in VA indicators based on non-test outcomes, with Jackson et al. (2020); Beuermann et al. (2022) 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 of the long-term consequences of differences in school quality, and also limited evidence on the validity of VA models outside North America.[2]

In this paper I study the VA of Norwegian compulsory schools (up to and including Year 10) 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. (2014a) use to estimate teacher VA. Detailed population-level administrative data allow me to construct measures of student background that enable me to estimate credible VA models, observe student outcomes into the students' early 30s, and track students that change schools.

I find persistent differences in VA across 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. (2014a). The relationships between VA and long-term student outcomes are mostly as strong as or stronger

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

[2] While most studies are from the US Beuermann et al. (2022) study causal effects of schools in Trinidad and Tobago.

than the corresponding cross-sectional student-level relationships between the students' grades and long-term outcomes. That is, the predicted gain from attending a high-VA school is for most long-term 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 small between-school differences,[3] 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 point difference in labor market participation and a 1.9 percent earnings difference (given positive earnings) at around age 32.

While previous US studies have found that VA estimates controlling for students' previous achievement produce unbiased estimates, lack of test data makes this approach unfeasible 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 for contemporaneously observed family background alone may provide informative VA estimates, at least in some contexts. This enables estimation of VA in school systems without sufficient data for historical standardized tests, like Norway, and makes it possible to study VA before the first tests are available, e.g. for the first years of primary school.

For recent cohorts I also observe standardized tests throughout the years of compulsory schooling. I give a brief account of how studying different periods and using different sets of controls impacts the interpretation and comparison of VA indicators. The results of investigating different VA measures for different stages of compulsory schooling provide a consistent picture of the impacts of pre-existing skills and schools' VA, and show that VA towards the end of compulsory schooling matters most for exam scores and completion of upper secondary school.

My main analysis does not distinguish between a school and the students at that school. Furthermore, most Norwegian students attend their neighborhood school. Thus, while the VA estimator for a given year only depends on the outcomes of students of other years, there may be persistent differences between schools in unobserved student characteristics if such differences exist across neighborhoods. To address this potential concern, 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

[3] 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.

to what we would expect from the VA estimated for 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 for concern with respect to the validity of the quasi-experiments. However, I find no indication based on observed student characteristics of movers sorting to high-VA schools. This suggests that while moving may be non-random, the VAs of movers' new schools are random (conditional on observed student characteristics), and thus that studying the movers constitutes a valid quasi-experiment for the effect of school VA on this group.

Studying movers also allows me to investigate transitory VA, i.e., year-to-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 impacts the outcomes of incoming students. Transitory VA is strongly associated with the outcomes of movers, but, unlike persistent VA, is not forecast-unbiased. This suggests that transitory VA in part reflects 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-year 10 grades given by the students' teachers. While these VA estimates are forecast-unbiased with respect to 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 level of the individual student. This likely indicates that classroom teachers are better able to observe students' performance 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 average performance than average exam scores. This interpretation is supported by the quasi-experiments, where I find that the exam scores of students who move do benefit from transitory exam score VA at the incoming school, while their teacher grades benefit much less.

The weaker association of teacher grade VA with long-term outcomes also provides information about the mechanisms through which school quality impacts later outcomes. Admission to upper secondary schools is based on the lower secondary grade point average (GPA), given at the end of Year 10, so that a higher GPA will provide more educational opportunities. However, when the GPA is calculated, teacher grades are given approximately 10-20 times the weight of written exam scores.[4] Thus, the finding that exam score-VA indicators matter more for long-term outcomes than teacher grade-VA indicators suggests 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 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 demonstrates how it, at least in the current context, is possible to construct informative VA indicators even without data on lagged achievement or the mechanism by which students are assigned to schools. 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 the end of Year 10, 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 until it is possible to study VA using lagged achievement. However, it may also be useful to study the impact of early school quality, even if pre-school achievement is not recorded.

Second, I study VA throughout the years of compulsory schooling. Previous studies typically study VA during a single year or some stage between tests.[5] However, VA estimates that do not require lagged achievement data allow us to study the whole compulsory schooling period and to study the effect of school quality at different stages, as Carneiro et al. (2021) do in their study of the timing of parental earnings. I find that late school quality matters most for later outcomes.

Third, I distinguish between persistent and transitory VA. Persistent VA, as estimated by Chetty et al. (2014a), 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 outcomes of movers

[4] The relative weight varies with the schooling year, with less weight on exams in earlier years, and to a lesser extent among students within a particular year.

[5] For example, Jackson et al. (2020) study value-added of secondary schools and control for end-of-primary (grade five) test scores, while Angrist et al. (2020) study value-added of middle shools and high schools.

strongly suggests that transitory VA reflects both within-school differences across 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 movers, we can conclude that instruction quality has both 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 the schools' effects on test scores. As teacher grades arguably reflect a broader set of skills and competencies, 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. 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, which is similar to what Chetty et al. (2011) do using Project STAR. Our study uses general variation in school quality, suggesting that the (implied) effect of learning on long-term outcomes may be generally relevant (as opposed e.g. to specific interventions, which 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 of the quasi-experiments. The final section provides a conclusion.

2 Institutional setting and data

2.1 Compulsory education in Norway

Compulsory education in Norway lasts for 10 years and is divided into primary (years 1-7) and lower secondary (8-10). The school system is almost exclusively public, with less than 5 percent of students attending private schools for their compulsory schooling.[6] Students are assigned to a school by the municipality on the basis of residence, and most students attend their local school. In some cases, parents may have the option of choosing a school other than the neighborhood school, but this is subject to capacity.

In Norwegian schools teachers will often teach students in different years and tend to follow the same students within the major divisions (primary, lower secondary) of the school system. In the first years of compulsory schooling teachers tend to be generalists, teaching a class in all or most subjects, while later they will typically have a limited number of subjects in which they teach students from different classes.[7]

Since 2007, students in years five and eight take national standardized tests in literacy, numeracy, and English. Since 2010, students in year nine have taken the same tests in literacy and numeracy as the year eight students. These tests are taken early in the academic year, and are often considered exit scores from the previous year. At the end of compulsory schooling 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 moving to 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 on the basis of their GPA, and are not guaranteed to be allocated 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 schooling will be high stakes.

[6] Most private schools are funded by the government in much the same way as public schools. These schools are only allowed to charge limited tuition fees. For-profit schools do not receive funding; in order to operate a private school the school must represent a faith-based or pedagogical alternative to the public schools. Less than 0.5 percent of students attend international schools that are not funded by the government.

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

While almost all students enroll in upper secondary education (about 98 percent enroll directly after finishing their compulsory schooling), drop-out and delayed completion of upper secondary school is considered a serious problem. The nominal duration of upper secondary is 3-4 years, but only about 75 percent of students 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 schooling (year 10) grades for the entire student population completing this schooling 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 the year in which they complete compulsory schooling. Thus, while exam scores and teachers grades are available for all cohorts 2002-2019, the 2010 cohort is the first for which the year eight test exists, and the 2012 (2013) cohort the first for which I observe the year nine (5) test. Within the cohorts for which tests are observable, few students have missing values (5-10 percent for each outcome, except for the year five 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 a mean of zero and a standard deviation of one within each cohort.

Students are linked to parents to permit the construction of 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, of students with at least one parent with higher education and shares of students who are immigrants or Norwegian-born with two immigrant parents. The share of highly-educated parents has increased steadily, from about 40 percent for the 2002 graduates to 54 percent for the 2019 graduates. The share of immigrant students increased before decreasing again, and is 7 percent 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 school, completed years of schooling and labor earnings. Post-compulsory schooling outcomes are measured up to or in 2019 (except completion of upper secondary school, which is also observed in 2020), i.e. 17 years after the first cohort completes compulsory schooling, 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 shown in Figure A4. The only outcome strictly limited by data availability is "on-time" upper secondary school completion, which is measured five years after completion of lower secondary school,[8] and is thus is observable for cohorts completing compulsory schooling in 2015 or earlier.

3 Empirical approach

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

The 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. The primary school results z_{is}^0 depend on the quality of primary school Q_s^0, student characteristics x_{is}, and an idiosyncratic error term:

$$z_{is}^0 = Q_s^0 + x_{is}\beta^0 + \epsilon_{is}^0 \tag{1}$$

Allowing for some persistence in results from primary to lower secondary, captured by the coefficient λ, results in lower secondary school z_{is}^1 can be expressed as a function of previous results, lower secondary school quality, and student background;[9]

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

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

I assume that x_{is} captures all sources of student-level persistence of 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. Given these assumptions, reorganizing (2), the difference between observed lower secondary school results and the results expected in light of the students' backgrounds and previous results reflects lower secondary school quality:

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

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

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

Eq. (3) is the traditional VA measure of school quality used in 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 school, where primary school quality is weighted by its persistence when determining results:

$$Q_s^{av} = \lambda Q_s^0 + Q_s^1 = E[z_{is}^1 - x_{is}(\lambda\beta^0 + \beta^1)] \tag{4}$$

3.1 Estimating school quality

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

$$z_{ist} = x_{ist}\beta + \epsilon_{ist} \tag{5}$$

Here, z_{ist} represent the results (typically exam or test scores) of student i completing at school s at time t.[10]

The vector of covariates (x_{ist}) will always include a cubic in the 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 scores on a set of dummies for gender*immigration status (native, immigrant, immigrant parents)*socioeconomic status (five categories based on parental education) and the combination of the levels of parents' highest completed educations, and obtain the predicted exam score for each student.[11] Other than the socioeconomic index and and the school-level average socioeconomic index, the set of controls always includes compulsory schooling completion year. Some specifications also include a cubic in the year eight test score (average of available tests), as well as the school mean for the average year eight test score.

As illustrated by equations (3) and (4), whether I control for previous results or not changes the interpretation of the VA indicators. Controlling for primary school results

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

[11] When constructing 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.

yields 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 schooling.

By 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. (2014a), I estimate persistent value-added by means of 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 autocorrelation vector, that may depend flexibly on time difference and thus captures persistence in school results. In contrast to Chetty et al. (2014a) I find that the correlations arefairly stable, almost irrespective of time difference, at 0.2-0.3 (lower when controlling for previous test scores). This is similar to the long-term correlation of Chetty et al. (2014a), but smaller than the short-term correlations. A likely explanation is that Chetty et al. (2014a) study teacher quality, which may be more persistent in the short term.[12] School quality, on the other hand, will change as different cohorts are taught by different teachers. However, although school quality varies more from year to year, there is still a stable component to it, reflecting some shared aspects of the school other than individual teachers. Because of this stability of the autocorrelation 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. (2014a), but with shorter lags adapted to the stable correlations).

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

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

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

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 the average value-added of school s for cohort t over and above the persistent quality, and will reflect the 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. (2014b); Deming et al. (2014); Angrist et al. (2016)), we expect $\hat{\mu}_{st}$ to be reflected in the results of a student entering school s and completing schooling 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, in the absence of peer effects, will not affect a randomly placed student).

3.2 Evaluating the effects of school quality

I next study associations between estimated persistent VA and short-term (exams, teachers grades) and long-term (further education, earnings) outcomes. The general regression equation relating each outcome y_{ist} of a student i completing at school s at time t 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. (2014a) and denote the VA indicators as (forecast) unbiased if γ=1, i.e. if the indicators on average forecast outcomes without error.

$\hat{Q}_{st}$, and also $\hat{\eta}_{st}$, will depend on the residuals ϵ_{ist} of students completing compulsory schooling at 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 $\boldsymbol{Q}_{s,-t}$, which is related to ϵ_{ist} only through persistent school differences. Interpreting unexplained persistent result differences between schools as reflecting school quality thus implies 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 may also be persistent differences between schools that do not reflect 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 because of residential sorting combined with neighborhood schools, for example, and may 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 in 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 makes it possible to estimate the effect of school quality on long-term outcomes. I will discuss the validity of the quasi-experiments further when presenting the results.

Assuming 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 that do not change 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 the school they attended or were assigned to).

4 Persistent 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 completing compulsory schooling from 2004-2019, 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 32-33. I construct VA indicators for three outcomes: end-of-compulsory written exam scores, oral exam

Table 1: Standard deviations of year 10 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. 2004-2019 completion cohorts.

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 cohorts at 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 scores than for the other outcomes. The VA indicators have standard deviations around 0.09-0.13, higher for teacher grades than for the exam scores. This is a smaller dispersion of VA than found in studies from other countries. Angrist et al. (2020) find standard deviation around 0.20 for NYC middle schools, and Beuermann et al. (2022) find a standard deviation of 0.45 among secondary schools in Trinidad and Tobago. The reduced dispersion of the VA indicators compared to the school-by-year averages reflects both averaging over cohorts and adjustment for between-school and between-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 skills and competencies that are restricted to exam scores, or the extent to which school value-added also predicts 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 assist in the 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 the sample. I find a coefficient of close to 1, and although slightly attenuated, not significantly different from one. In other words, the exam-based VA indicators forecast average out-of-sample exam scores 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 school and completion of upper secondary school within five years, are both more strongly associated with the exam-based indicator than with own exam score.

In the next columns I show the results of similar analyses of longer-term outcomes related to earnings and labor market participation. A potential challenge in studying these outcomes is that more academically successful students stay longer in school, which may influence measurement of the outcomes. In column (6) we see that this is indeed the case. 9 percent of all students are still in education 16-17 years after completing compulsory schooling, and students with higher exam scores more often. However, the VA indicator is negatively related to whether the student is still in education. The VA indicator is also related to labor market outcomes. The share employed[13] 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 point difference in labor market par-

[13]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) Compl. upper sec. school	(5) Years schooling	(6) In education	(7) Employed	(8) NEET	(9) Earnings (NOK 100k)	(10) Log earnings
$\hat{\mu}$ written exam	0.919**	0.655**	0.762**	0.309**	0.813**	-0.026**	0.077**	-0.058**	1.599**	0.199**
	(0.057)	(0.058)	(0.066)	(0.024)	(0.159)	(0.012)	(0.015)	(0.014)	(0.155)	(0.023)
$\hat{\mu}$ oral exam	0.677**	0.969**	1.026**	0.143**	0.610**	-0.004	0.041**	-0.037**	1.267**	0.175**
	(0.061)	(0.056)	(0.062)	(0.026)	(0.197)	(0.011)	(0.014)	(0.012)	(0.161)	(0.024)
$\hat{\mu}$ teacher	0.437**	0.589**	1.007**	0.151**	0.508**	-0.006	0.052**	-0.047**	1.097**	0.129**
	(0.045)	(0.044)	(0.042)	(0.019)	(0.150)	(0.009)	(0.010)	(0.009)	(0.118)	(0.017)
Written exam score	1.000	0.514**	0.643**	0.153**	0.944**	0.013**	0.035**	-0.035**	0.644**	0.087**
	(.)	(0.004)	(0.004)	(0.002)	(0.015)	(0.001)	(0.001)	(0.001)	(0.014)	(0.002)
Oral exam score	0.492**	1.000	0.606**	0.150**	0.881**	0.012**	0.037**	-0.035**	0.631**	0.083**
	(0.004)	(.)	(0.003)	(0.002)	(0.015)	(0.001)	(0.001)	(0.001)	(0.012)	(0.002)
Teacher grade	0.708**	0.698**	1.000	0.236**	1.309**	0.013**	0.059**	-0.056**	0.866**	0.110**
	(0.004)	(0.003)	(.)	(0.002)	(0.015)	(0.001)	(0.002)	(0.002)	(0.014)	(0.002)
N	83372	83372	83145	82723	81154	83372	83372	83372	81135	72504
# clusters	2026	2026	2025	2026	2025	2026	2026	2026	2025	2012
$\bar{y}$	0.015	0.015	0.037	0.714	13.909	0.089	0.870	0.114	5.069	1.642

Note: Each cell is a separate regression of the outcome on a VA indicator or exam/teacher grade on the 2002 and 2003 Year 10 cohorts. Outcomes (1)-(3) are from the end of compulsory schooling (Year 10) and (4) is observed five years later. Outcomes (5)-(10) are observed in 2019, i.e. 16-17 years after completion of compulsory schooling, when subjects are around age 32-33. (5) is nominal duration of highest completed education (in years, including compulsory schooling); (6) is a dummy for whether the person is in education in 2019; (7) is an earnings-based employment measure (earnings > G, approx USD 10 000); (8) is a dummy for not in employment, education or training; (9) is annual labor earnings and (10) is log annual earnings. The indicators are constructed from the 2004-2019 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%

ticipation 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 compared with the individual-level cross-sectional associations, and the differences in secondary school completion and participation are also substantial compared with 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 across 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. Nor are indicators based on teacher grades 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 most differences are not 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.

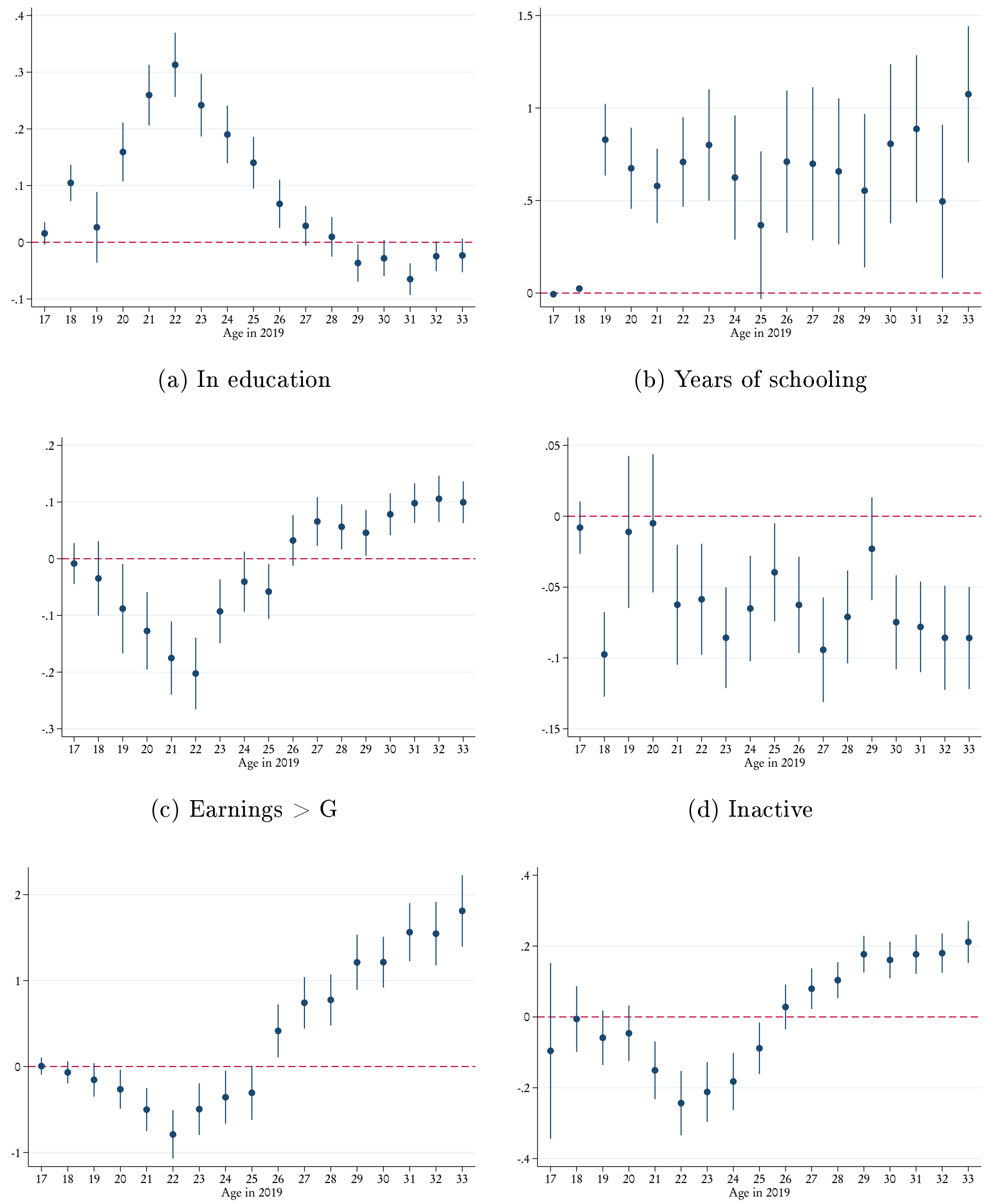
Thus, indicators constructed from oral exam scores or teacher grades also capture persistent differences across schools. These indicators are also predictive of later outcomes, but less so than indicators based on written exam scores, in particular teacher grade VA. This is despite the fact that student-level 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 by 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 of student backgrounds, see 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 schols on the basis of 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 municipalities, which may also differ in other ways, 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 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 Year 10, cf. Section 3.

Panel (a) of Figure 1 shows the association between VA and being in education. Almost all Norwegian students start upper secondary school after completing their compulsory schooling, so 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 cohorts' early 20s, and is reversed at 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



(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 2004-2019 cohorts. Standard errors adjusted for school-by-cohort clustering.

ages when more young people are in education. However, students from high-VA schools have persistently higher labor market participation from their late 20s onward. Nor does the association of VA with education and employment 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 earning gains from completing compulsory schooling at a high-VA school stabilize from around age 30.

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

As in the previous section, when studying longer-term educational and labor market outcomes I am restricted to using cohorts from before the introduction of the year five and eight tests. In the current section I will use more recent cohorts, for whom standardized year five and eight tests are available, to study indicators using different sets of control variables. I will also investigate transitory VA (see discussion in Section 3), and compare its dispersion with that of persistent VA. As in the previous section, I will estimate VA indicators for a set of cohorts, in this case the 2015-2019 completion cohorts, and reserve earlier cohorts for testing the out-of-sample performance of the indicators.[14]

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

[14] The 2013 cohort is the first for whom year 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 year five and eight test scores and VA indicators estimated from later cohorts.

In Table 3 I present SDs of test scores, school-by-year means and the different VA indicators, in the same way as for the VA indicators constructed from the 2005-2019 cohorts in Table 1. In Figures A7-A9 in the Appendix I show the corresponding distributions. All outcomes are standardized at student level, and thus have comparable scales. Furthermore, all outcomes have school-by-year means close to 0.3, like those for written exam scores in Table 1. Restricting the sample to more recent cohorts reduces the standard deviation of persistent exam score VA somewhat, from 0.093 to 0.083. The dispersion of VA indicators estimated from the year eight and year five test scores are greater, with standard deviations of 0.114 and 0.134 student-level standard deviations respectively. There are approximately twice as many primary schools with year five VA indicators as lower secondary schools with exam score indicators, which probably contributes to greater school-level dispersion. However, the year eight scores are associated with lower secondary schools, even though the outcome is essentially end-of-primary proficiency. Thus, the aggregation of the year 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.[15] The lower part of Table 3 shows results for indicators controlling for previous test scores. VA for exam scores from a specification controlling for year eight test scores has slightly lower dispersion than VA form a specification controlling only for family background. The dispersion of VA for year eight tests when controlling for the previous test is very similar to 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 either by student characteristics or by 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 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 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, in the same way as in Table 2. I regress student-level exam scores of students completing compulsory schooling in 2013-2014 on VA indicators constructed from

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

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

	Written exam score	Year eight test	Year five 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 following rows. 2015-2019 completion cohorts.

the 2015-2019 cohorts, controlling for the same control variables as were 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 determine the interpretation of the estimated indicators. Furthermore, when regressing exam scores on school quality, controls for student background need to be consistent with the indicators. In Table 4 I break down the total contributions from school quality and student 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 pre-year five, year five to eight and year eight to the end of compulsory schooling, as well as students' (pre-school) background.[16] On average, VA from year eight to exam scores predict exam without bias. The estimated coefficient on the value-added indicator is precisely estimated and not significantly different from one. Value-added in year five to eight and pre year five also predict exam scores, conditional on exam value-added and student background. As the year five and eight value-added measures are in units of test scores while the outcome is exam score, we cannot strictly speaking talk about "unbiased", as we did for exam value-added. However, as each of these outcomes is measured in terms of student-level standard deviations, the scales are

[16] Student background is partly decided at birth (sex and immigrant background) and partly measured at age 16 (parents' education). However, although parents' formal education may change during a student's childhood, this is relatively rare. Furthermore, in the context of their children's school performance, parents' education is also a proxy for fixed parent characteristics that correlated with education. While the background variables themselves reflect pre-school characteristics, the relationships with exam scores may change over time. As illustrated by eqs. (3) and (4) the student background coefficients will represent the total association between background and school performance.

Table 4: Predicting exam scores of 2013-201 4Year 10s

	(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)		
Year eight value-added	0.646**		0.635**		0.485**	
	(0.065)		(0.068)		(0.045)	
Year five value-added	0.238**					0.239**
	(0.031)					(0.035)
Controls (cubic + school mean):						
Year eight score		Yes				
Year five 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 school level. Significant at * 10%, ** 5%

directly comparable. From Table 3 we also know that the dispersion of the indicators is similar. Thus, year eight and to an even greater extent year five value-added are less strongly associated with exam scores than exam score value-added, as a one SD difference in value-added is 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 year eight 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 from either the coefficient in column (1) or from one. Similarly in column (3), when I control for year five score and disregard school quality in lower secondary and pre-year five, the coefficient on year eight value-added is not significantly different from the coefficient in column (1).

Columns (4)-(6) similarly regress exam scores on indicators constructed controlling only for socioeconomic background. Consistent with how the indicators are constructed, these specifications control only for student background, and not for previous test scores. All the associations between the indicators and exam scores are very similar to those

Table 5: Predicting completion of upper secondary schooling for 2013-2014 Year 10s

	(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)		
Year eight value-added	0.108**		0.100**		0.109**	
	(0.023)		(0.022)		(0.015)	
Year five value-added	0.065**					0.065**
	(0.011)					(0.011)
Controls (cubic + school mean):						
Year eight score		Yes				
Year five score			Yes			
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

Note: Outcome is completion of upper secondary schooling 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 school level. Significant at * 10%, ** 5%

in column (1), and again the exam value-added indicator predicts exam scores without bias.

For the 2013-2014 Year 10s I have data both on test scores and on completion of upper secondary school. In Table 5 I show how upper secondary school completion is related to VA for different stages of compulsory schooling, as I did for exam scores in Table 4. VA indicators for exam score, year eight test score and year five test score all predict completion, for each permutation of VA indicators and controls. VA later in compulsory schooling is more strongly related to completion of upper secondary school than VA from earlier stages, as for exam scores in Table 4, However, the gradient is less pronounced than in Table 4.

In Figure 2 I present the results of regressing exam scores on value-added, as in columns (2) and (4) of Table 4, for each of the years 2010--2019. As in Table 4 VA is based on the 2015-2019 cohorts. Unsurprisingly, for each year of the estimation period the coefficients are close to 1. Before the estimation period, the coefficients are close to 1 in the two years immediately preceding the estimation period, but are less than one for greater time differences. However, three to five years before the estimation period, the

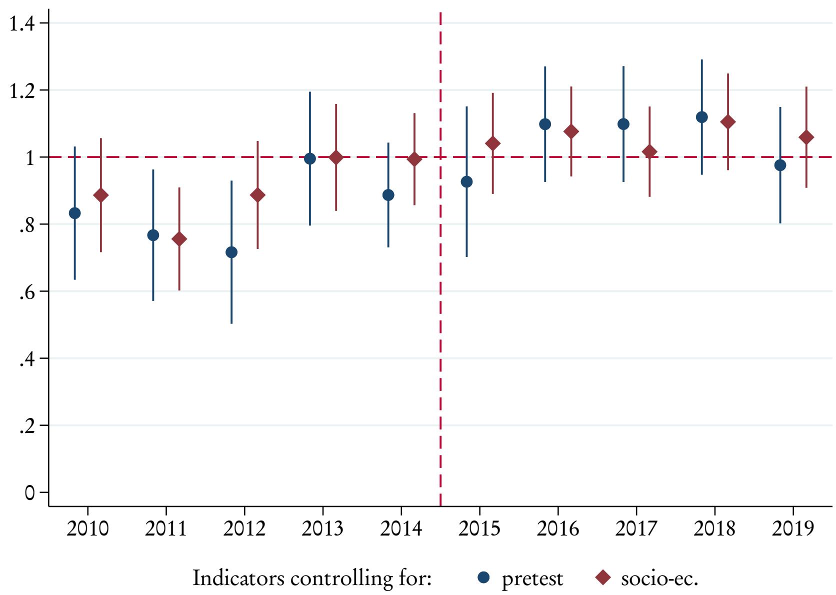


Figure 2: Exam score and estimated school quality by cohort
Note: The graph shows the estimated relationship between exam score and VA indicators, as in columns (2) and (4) of Table 4. VA indicators are based on the 2015-2019 cohorts.

coefficients are still 0.7-0.9, and the confidence intervals overlap with the coefficients of the previous years. Thus, the VA captures 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 actually reflect school quality, or whether they reflect other unobserved differences.

An important concern when trying to disentangle school quality from student characteristics is related to the potential for systematic sorting of students to different schools. If such sorting is present, students attending the same school may share unobserved characteristics that may confound the analysis. In Norwegian compulsory schooling, 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 which neighborhood their families choose to live in and thus of characteristics correlated with this choice.

As there is no assignment to schools by lottery in Norwegian compulsory education, I rather rely on three different quasi-experiments, where students change their actual or predicted school in different ways. The first uses students who are observed in different schools, the second uses students that move between municipalities, while in the last the assignment of neighborhoods to specific schools changes. In each of these situations I estimate VA based on students/neighborhoods not changing schools (stayers) and see how the outcomes of the students/neighborhoods that do change schools (movers) are related to these estimates. It is probably not random which students or neighborhoods change schools. However, I discuss whether the VA of the new school can be conditionally considered as good-as-random, and thus whether the we can estimate the effects of the new (and potentially the old) school on the movers.

6.1 Students observed in different schools

While Norway lacks a central registry of which school compulsory schooling students attend, since 2007 students' school assignment can be observed when the students sit the standardized tests in years five and eight, and since 2010 also in year nine. This enables us to observe the students' school assignment several times throughout compulsory schooling. In particular, the tests in year eight and nine enable identification of students who change school early in lower secondary school. Some students are observed at one school for the year eight test and a different school in year nine; I 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 complete their compulsory schooling at the same school as they sit the year eight test (92 percent of the students). I then 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 provide a consistent estimate of the effect of the value-added of the new school.

In Table 6 I present the results from regressing the outcomes of changers on the persistent ($\hat{\mu}$) and transitory ($\hat{\eta}$) VA of the old (year eight) and new (year nine) school, constructed using non-changers. All regressions also include controls for average characteristics of the student cohorts in the old and new schools.

The upper panel presents results based on indicators controlling for lagged achievement, so that VA is VA for lower secondary school, conditional on achievement at the start of lower secondary school, as in eq. (3). In the lower panel I use indicators controlling only for family background, so that VA is the combined VA of primary and lower secondary school, as in eq. (4). In column (1) I present the results of 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 1 is to be expected, as the students change school some time between early year eight and early year nine, and thus do not spend all of lower secondary school in their new school. Furthermore, of those changing school at least once, about 20 percent change again before completing compulsory schooling. However, while the coefficient is less than 1, it is not significantly so. By way of comparison, the coefficient on the persistent VA of the old school is 0.28.

Because school changers and non-changers used to estimate transitory VA are separate but concurrent groups, I can also study how transitory VA is related to the outcomes of students not used to construct the indicators. We see that the 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 probably captures a wide range of causes of differences in results, such as the 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 are often taught by different teachers, this likely reflects, at least in part, within-school differences in teacher VA; see Chetty et al. (2014a). 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, so 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, the 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)

Table 6: Exam score VA and outcomes for school-changers

	(1) Written exam	(2) Teacher grade	(3) Completed upper secondary school	(4) NEET	(5) Control (pretest/ index)
Indicators controlling for pretest					
$\hat{\mu}^{Old}$	0.282**	0.269**	0.204*	-0.058	0.143
	(0.116)	(0.123)	(0.116)	(0.103)	(0.142)
$\hat{\mu}^{New}$	0.879**	0.504**	0.242**	-0.031	-0.224
	(0.122)	(0.132)	(0.106)	(0.100)	(0.138)
$\hat{\eta}^{Old}$	0.119**	0.067	0.060	-0.044	0.022
	(0.046)	(0.046)	(0.037)	(0.039)	(0.050)
$\hat{\eta}^{New}$	0.322**	-0.007	0.017	-0.015	-0.013
	(0.050)	(0.044)	(0.036)	(0.035)	(0.048)
Indicators controlling for family background					
$\hat{\mu}^{Old}$	0.269**	0.419**	0.199**	-0.065	0.039
	(0.089)	(0.103)	(0.072)	(0.066)	(0.059)
$\hat{\mu}^{New}$	0.756**	0.547**	0.208**	-0.115*	0.017
	(0.090)	(0.105)	(0.072)	(0.065)	(0.056)
$\hat{\eta}^{Old}$	0.085*	0.088*	0.051	-0.057	-0.007
	(0.047)	(0.052)	(0.035)	(0.036)	(0.027)
$\hat{\eta}^{New}$	0.325**	0.076	0.045	-0.020	0.006
	(0.046)	(0.048)	(0.035)	(0.034)	(0.025)
Student controls	Yes	Yes	Yes	Yes	*
N students	8013	7827	4236	3229	8013
N clusters	935	932	842	798	935

Note: Each column*panel is a separate regression of an outcome on persistent ($\hat{\mu}$) and transitory ($\hat{\eta}$) VA of the year eight and year nine schools. The sample consists of all students with recorded year eight and nine tests (compulsory schooling completion cohorts 2011-2019). Indicators are constructed from students in the same cohorts who don't change schools. Outcomes in columns (1)-(4) are the same as the corresponding outcomes in Table 2. Completion of upper secondary is observed only for students completers 2015 or earlier, NEET only for completers 2014 or earlier. The outcome in column (5) is the key individual control variable in the regressions in the same panel; the year eight test score in the upper panel and student background index in the lower panel. All regressions in columns (1)-(4) control for student background by cubic in background index, in the upper panel also for cubic in year eight test score. The regression in the upper panel of column (5) controls for student background, but not test score, the regression in the lower panel for neither. All regressions include school-level controls for the year eight and nine schools (school*year average student background, and also test score in the upper panel). Standard errors are clustered at the year eight-school. Significant at * 10%, ** 5%.

is that there is no relationship between the transitory VA of the new school and teacher grades. The significant effect of transitory VA on exam scores in column (1) indicates that transitory VA affects 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 an indication 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 year eight test score is related to the VA measures. While there are strong (and possibly causal) positive relationships between the persistent and transitory VA of the new school and exam scores, there are insignificant negative relationships between VA and the pre-determined year eight test scores. A lack of a significant relationship is to be expected, given our knowledge of the context. Historically, data on school quality has not been easily available in Norway. Average end of compulsory school grades have been available since 2002, but data on VA has not been available, and data on transitory VA is not even forecastable. 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 for 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 year eight 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 year eight 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 while controlling for student background (and not for lagged achievement) is related to exam scores, and not very strongly so. This suggests that VA indicators based on teacher grades mostly measure differences in grading practices, rather than differences in school quality.

6.2 Students that move between municipalities

For older cohorts that completed grade eight before 2008 we are not able to observe school assignment until they completed Year 10. 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.[17] The student cohort completing compulsory schooling in 2017 is spread across 11,000 neighborhoods, with 1-88 students in each (the average is 6 students). I find the modal school for students in the neighborhood for each neighborhood and cohort of compulsory school students. As data on school assignment only is available at the end of compulsory schooling, this will provide a predicted lower secondary school. While some students attend the same school throughout compulsory schooling, many schools are only primary or lower secondary schools, and many students change school at the transition from primary to lower secondary. Thus, I cannot predict the primary school attended. Rather, I study value-added associated with lower secondary schools, acknowledging that this may in part stem from the contribution of the primary schools previously attended; see the 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 schooling, while movers are students that move between municipalities at least once during compulsory schooling. I use the never-movers to estimate value-added based on exam scores and teacher grades, and construct VA for each neighborhood and cohort as the VA of the modal school. As test score data is not available for these cohorts, I only control for the socioeconomic index.

Table 7 shows the relationships between movers' outcomes and the persistent and

[17] 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 populations in 2017 ranging from 1-6000 (the average population is 379). The units are described as *"small, stable geographical units* which may form a flexible basis for working with and presenting regional statistics (...) *geographically coherent* areas (...) *homogeneous* with respect to nature and basis for economic activities, conditions for communications, and structure of buildings" (my emphasis).

transitory VA associated with the movers' neighborhood at the start of compulsory schooling 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 complete at the modal school of their neighborhood after their first move.[18] Thus, if the VA of the school actually attended by the movers is uncorrelated with the VA of their predicted school when these schools are not one and 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 1. As 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)-(6) I show how persistent and transitory VA are related to teacher grades and longer-term outcomes. Both persistent and transitory VA of the new school are significantly related to teacher grades, completion of upper secondary 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, the 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)-(6) the coefficients on transitory VA are about 2/5 of the coefficient on persistent VA. The VA of the old neighborhood is generally about as strongly related to the outcomes in columns (2)-(6) as the VA of the new neighborhood.

For the coefficients on the new neighborhood VA to be informative about the 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 the persistent and the transitory

[18] 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 complete at the school at which they are expected to complete their compulsory schooling, 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 the neighborhood in each year. However, this requires deciding how to weigh the 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, the inclusion of later schools complicates the interpretation of the coefficients on the second school.

VA of the new neighborhood are unrelated to student background; the coefficients are insignificant and close to zero (recall that the background index is the predicted exam score, and thus has the same scale as the exam score). There are however significant associations between the background and both the persistent and the transitory VA of the old neighborhood. The association with the background index is also close to the observed association with the exam score VA of the old school, in particular for transitory VA.

In the lower panel I present the results of 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 the persistent and the transitory VA of the old neighborhood are still significantly related to student background.

The movers are not involved in estimating the VA of either the old or the new school. The association between old-school VA and student characteristics indicates that 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 exam scores than persistent exam score VA. While transitory teacher grade VA predicts exam scores, it only weakly predicts teacher grades, similar to exam score VA. As in the previous sub-section, this likely reflects relative grading. However, both the persistent and the transitory teacher grade VA indicators 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 study changes in the schools' catchment areas,

Table 7: Exam score VA and mover outcomes

	(1) Written exam	(2) Teacher grade	(3) Complete upper secondary school	(4) NEET	(5) Earnings	(6) Log earnings	(7) Background index
$\hat{\mu}^{Old}$	0.263**	0.397**	0.144**	-0.069**	0.351**	0.092**	0.054**
	(0.037)	(0.040)	(0.021)	(0.017)	(0.179)	(0.034)	(0.023)
$\hat{\mu}^{New}$	0.601**	0.378**	0.140**	-0.023	0.635**	0.058*	-0.013
	(0.034)	(0.037)	(0.020)	(0.016)	(0.161)	(0.032)	(0.017)
$\hat{\eta}^{Old}$	0.050**	0.045**	0.019**	-0.014**	0.088	-0.005	0.037**
	(0.015)	(0.015)	(0.008)	(0.007)	(0.068)	(0.013)	(0.007)
$\hat{\eta}^{New}$	0.341**	0.085**	0.043**	-0.007	0.144**	0.025*	0.002
	(0.015)	(0.015)	(0.009)	(0.007)	(0.067)	(0.013)	(0.007)
With neighborhood fixed effects:							
$\hat{\mu}^{Old}$	0.259**	0.288**	0.064	-0.063	-1.439**	-0.258*	0.120**
	(0.090)	(0.092)	(0.054)	(0.049)	(0.684)	(0.138)	(0.040)
$\hat{\mu}^{New}$	0.592**	0.396**	0.142**	-0.030*	0.528**	0.016	0.006
	(0.038)	(0.040)	(0.021)	(0.018)	(0.194)	(0.039)	(0.018)
$\hat{\eta}^{Old}$	0.041**	0.028*	0.005	-0.013	-0.046	-0.044**	0.026**
	(0.017)	(0.017)	(0.010)	(0.008)	(0.093)	(0.019)	(0.008)
$\hat{\eta}^{New}$	0.342**	0.065**	0.044**	-0.003	0.089	0.025	0.003
	(0.016)	(0.016)	(0.009)	(0.008)	(0.076)	(0.015)	(0.007)
N students	95922	98294	83092	78148	38510	31209	104805
N clusters	10441	10478	10153	10030	8182	7617	10589

Note: The sample consists of students moving during compulsory schooling. Outcomes in columns (1)-(6) are the same as in Table 2. Completion of upper secondary is observed only for students completers 2015 or earlier, NEET only for completers 2014 or earlier and earnings only for completers 2008 or earlier. $\hat{\mu}^{Old}$ 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 indicators 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 except (7) 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%

which arguably are exogenous to the students. As very limited data exist on school catchment areas, I infer these from the students' neighborhoods, as in the previous subsection. To find neighborhoods that change school assignment, I identify neighborhoods in which students in each year prior to some year t overwhelmingly attend one school (meaning that at least 80 percent of the students in the neighborhood attend the school, and only considering neighborhoods with cohorts of at least four students) and then in t and all following years attend some different school. In a manner analogous to the quasi-experiments in the previous sections, I estimate value-added from the students in neighborhoods that do not change school assignment, and study whether these VA indicators predict outcomes for students in the neighborhoods that change 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 show the student-weighted distributions of the years of change and the difference between year 10 and the year of the change.

One challenge in interpreting the results of these analyses is that I do not observe the process leading up to and following the school change. I only observe that students from a given neighborhood attend one school before and another after a given year. This may 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 how long students have been attending that school. For students completing compulsory schooling 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. However, as the circumstances concerning the change in catchment areas are unclear, I will disregard the first cohort completing compulsory schooling at the new school.

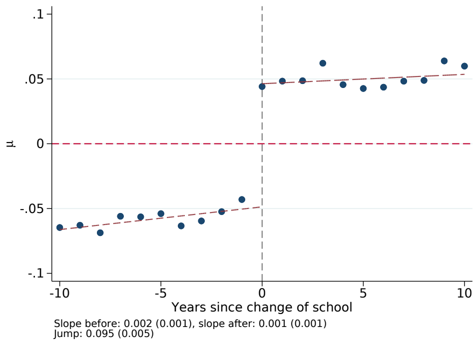
As I am able to follow neighborhoods and see how students' outcomes evolve over time, this natural experiment lends itself to an event study. In Figure 3 I show student outcomes in neighborhoods with an absolute change in predicted VA of at least 0.05 SD. Sub-figure (a) shows the average change in predicted VA. In all the sub-figures of Figure 3 outcomes are multiplied by the sign of the change in predicted VA, so on average they are expected to change from negative to positive. This is very clear for average predicted VA, which changes from -0.05 to 0.05, i.e. an average absolute change of 0.1 SD. Apart from in connection with the discontinuous change from old to new school, there is little evidence of VA trends. 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 0.011 SD, in the same direction as the change in predicted VA, but the change is not significant. Finally, sub-figure (d) 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, similarly as for persistent VA). This substantially reduces the dispersion of the annual averages. Residualized exam scores have an average change of 0.081 SD. This change is significantly different from zero and not significantly different from the change in predicted VA.

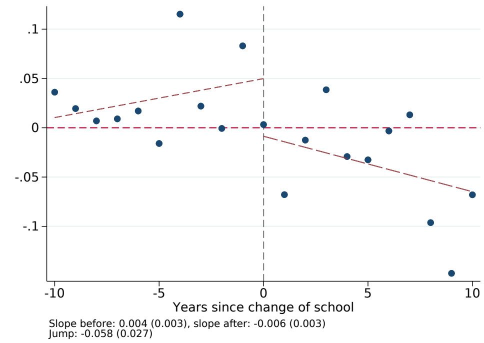
In Table 8 I study the relationship between the outcomes of students in neighborhoods that change school assignment and value-added estimated using students in never-changing units in a more parametric way, and include students in units whose predicted VA changes by less than 0.05. Column (1) shows how exam scores are related to the VA of the old and new school for students completing Year 10 at the new school (upper panel) and at the old school (lower panel). The exam scores of students completing Year 10 at 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 1. This is what we would expect if students have on average spent a substantial amount of time at the new school, but also, earlier, at the old school. Exam scores are strongly related to the transitory VA of the new school, and unrelated to the transitory VA of the old school. In contrast, the exam scores of students completing Year 10 at the old school are strongly related to both the transitory and the persistent VA of the old school, weakly related to the persistent VA (significant at the 10 percent level) and unrelated to the transitory VA of the new school.

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

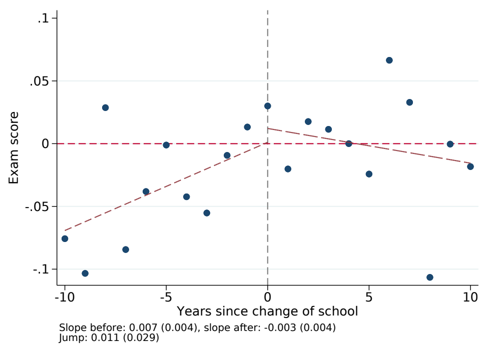
Completion of upper secondary school is related both to the VA of the old school and the VA of the new school for students completing lower secondary in either the old or the new school. This may reflect changes of school assignment also being concurrent with other changes that impact students in upper secondary school. However, while completion of upper secondary for student completing at the old school is impacted by the transitory VA of the old school, this is not the case for students completing their



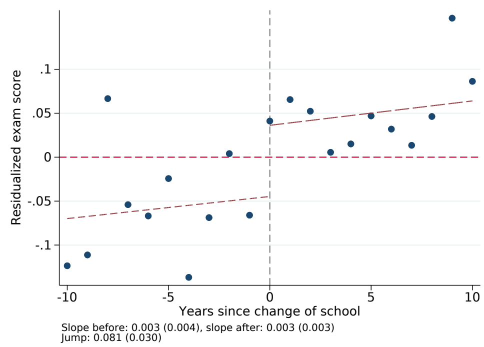
(a) Predicted persistent VA



(b) Predicted transitory VA



(c) Exam score



(d) Residualized exam score

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

Note: Sample is 38,759 students completing their compulsory schooling within 10 years of a change of predicted school that gives $|\Delta\hat{\mu}| > 0.05$. All outcomes are multiplied by $sign(\Delta\hat{\mu})$, such that VA and average outcomes are expected to change from negative to positive. For example, in sub-figure (a) the predicted persistent VA is $\tilde{\mu} = \hat{\mu} \cdot sign(\Delta\hat{\mu})$. Sub-figure (b) shows predicted transitory VA, sub-figure (c) shows observed exam scores, while sub-figure (d) 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 compulsory schooling at the new school. NEET is only related to persistent VA of the old school, and only for students completing at the new school. However, these estimates are not very precise. The estimated effect of the persistent VA of the new and old school for students completing at the new school and the persistent VA of the old school for students completing here are all neither significantly different nor significantly different from the corresponding estimate in Table 2.

In the last column in Table 8 I investigate how persistent and transitory VA is associated with student background. Unlike in Tables 6 and 7 there are significant associations, also between VA of the new school and the background of students completing at this school. However, as the background index has the same scale as the exam score, it is clear that the association between VA and background is much weaker than between VA and exam scores.

In Table A5 in the Appendix I repeat the analyses in Table 8 with controls for neighborhood fixed effects, thus controlling more flexibly for fixed characteristics of the neighborhood. This removes any significant associations between student background. This analysis also makes it very clear how exam scores are related to the persistent and transitory VA of the new (old) school for students that complete Year 10 at the new (school), with limited cross-effects from the other school. Teacher grades are significantly related to the persistent VA of the new school for students completing Year 10 at this school, and otherwise not related to VA. Longer-term outcomes are strongly related to persistent VA of the new school, but these results are very imprecise.

In Table A6 in the Appendix I repeat the analyses in Table 8 but with VA constructed from teacher grades. Persistent and transitory teacher grade-VA of the new (old) school matter for the teacher grades of students completing at the new (old) school, with no significant cross effects. Teacher grade-VA is much more weakly associated with exam scores, and also more weakly associated with longer-term outcomes than exam score-VA. This is consistent with the results in Tables 6 and 7 and 8, reinforcing the impression that teacher grade-VA to a large extent reflect grading practices.

7 Conclusion

Schools are a key instrument of policymakers for fostering skills and providing all children with opportunities. It thus of great relevance to identify schools that do this to a greater or lesser extent. In this paper, I study school quality in Norwegian compulsory schooling. Previous studies have found important differences in school VA in the US and Caribbean. I find persistent differences in VA also in Norway, with important consequences for

Table 8: Effect of change in exam VA due to change in predicted school

	(1) Exam score	(2) Teacher grade	(3) Complete upper secondary school	(4) NEET	(5) Background index
Students completing compulsory schooling at new school					
$\hat{\mu}^{New}$	0.916**	0.482**	0.120**	-0.041	0.126*
	(0.103)	(0.103)	(0.052)	(0.039)	(0.068)
$\hat{\mu}^{Old}$	0.473**	0.276**	0.147**	-0.093**	0.213**
	(0.120)	(0.128)	(0.063)	(0.040)	(0.072)
$\hat{\eta}^{New}$	0.150**	-0.003	0.006	0.004	0.024**
	(0.021)	(0.020)	(0.011)	(0.007)	(0.010)
$\hat{\eta}^{Old}$	0.013	0.058	0.008	-0.006	0.017
	(0.051)	(0.059)	(0.025)	(0.014)	(0.024)
Students completing compulsory schooling at old school					
$\hat{\mu}^{New}$	0.168*	0.230*	0.174**	0.007	-0.041
	(0.101)	(0.125)	(0.054)	(0.032)	(0.065)
$\hat{\mu}^{Old}$	0.797**	0.331**	0.119**	-0.034	0.036
	(0.103)	(0.127)	(0.057)	(0.031)	(0.061)
$\hat{\eta}^{New}$	-0.028	0.006	0.007	-0.015	-0.016
	(0.043)	(0.044)	(0.019)	(0.011)	(0.020)
$\hat{\eta}^{Old}$	0.268**	0.048**	0.024**	-0.007	-0.029**
	(0.021)	(0.021)	(0.010)	(0.006)	(0.010)
N students	66789	66160	54617	51633	66789
Nclusters	1212	1212	1206	1204	1212

Note: The sample consists of students living in a neighborhood (basic statistical unit) that has a change assigned school. Outcomes in columns (1)-(4) are the same as in Table 2. Completion of upper secondary is observed only for students completers 2015 or earlier, NEET only for completers 2014 or earlier. $\hat{\mu}^{Old}$ and $\hat{\eta}^{Old}$ are persistent and transitory VA (exam scores adjusted for student background) of the original modal lower secondary school in the students' neighborhood, while $\hat{\mu}^{New}$ and $\hat{\eta}^{New}$ are the VA indicators of the new modal school. The upper (lower) panel shows results for students completing lower secondary schooling after (before) the change in modal school. All columns control for cohort and neighborhood-average student background, all columns except (5) control for a cubic in the socioeconomic background index. Cluster (neighborhood)-robust standard errors in parentheses. Significant at * 10%, ** 5%

students' long-term outcomes.

Estimating persistent school quality as shrinkage-adjusted VA estimates by 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 labor market outcomes. Three quasi-experiments, in which students move or change school, or the association between neighborhood and school is changed, allow me to estimate value-added from a group of stayers, and investigate how outcomes for 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, there is no indication in any of the analyses that the identifying assumption, i.e. that changes in value-added are conditionally independent of student characteristics, is violated. I thus conclude that persistent VA measures are good measures of school quality.

Compared to 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 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 grade point averages are based mainly on teacher grades, this is also evidence that high-VA lower secondary schools impact long-term outcomes most 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 of and limited scope for bias in indicators controlling either for previous test scores only or for socioeconomic background only. 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 makes it possible to estimate school quality at early stages of primary school, where prior tests are usually not available, and to study the long-term outcomes for students for whom early test data are 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 evaluate different initiatives and policies, this valuation of school quality is important to enable better interpretation of the findings of such studies and prioritization of resources.

References

Angrist, J., P. Hull, P. Pathak, and C. Walters (2016). Interpreting tests of school vam validity. *American Economic Review 106*(5), 388–92.

Angrist, J., P. Hull, P. A. Pathak, and C. R. Walters (2020, December). Simple and credible value-added estimation using centralized school assignment. Working Paper 28241, National Bureau of Economic Research.

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

Beuermann, D., C. K. Jackson, L. Navarro-Sola, and F. Pardo (2022). What is a good school, and can parents tell? evidence on the multidimensionality of school output. *The Review of Economic Studies (forthcoming).*

Carneiro, P., I. L. García, K. G. Salvanes, and E. Tominey (2021). Intergenerational mobility and the timing of parental income. *Journal of Political Economy 129*(3), 757–788.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? evidence from project star *. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a, September). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

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

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

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

Falch, T., O. H. Nyhus, and B. Strøm (2014, 02). Performance of Young Adults: The Importance of Different Skills. *CESifo Economic Studies 60*(2), 435–462.

Fredriksson, P., B. Öckert, and H. Oosterbeek (2012). Long-term effects of class size. *The Quarterly Journal of Economics 128*(1), 249–285.

Fryer, Roland G, J. (2017). *The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments*, Volume 2, pp. 95–322. Amsterdam: North-Holland.

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

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non-test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jackson, C. K., S. C. Porter, J. Q. Easton, and S. Kiguel (2020, December). Who benefits from attending effective schools? examining heterogeneity in high school impacts. Working Paper 28194, National Bureau of Economic Research.

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

OECD (2006). *Education at a Glance 2006*.

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

Appendix

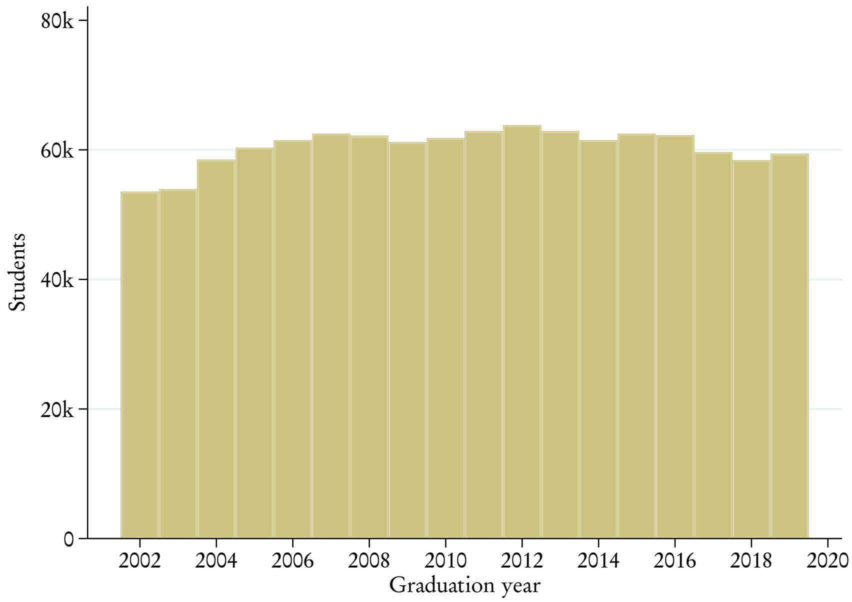


Figure A1: Cohort sizes

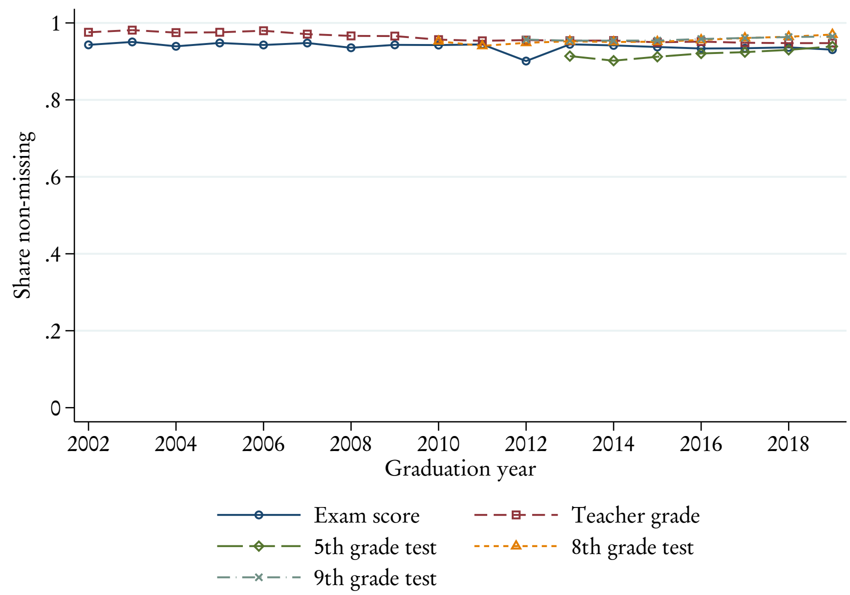


Figure A2: Share of students with non-missing exam scores, teacher grades and test scores by cohort

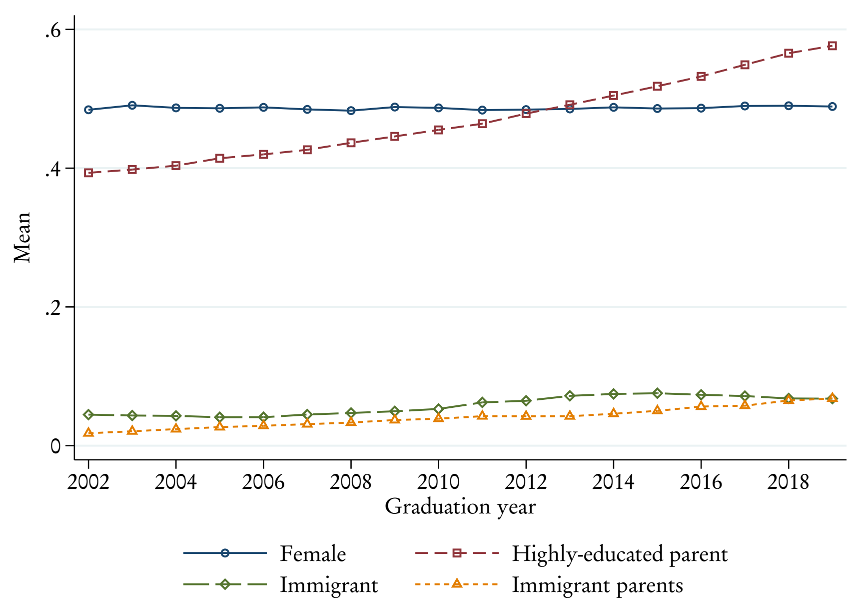


Figure A3: Mean background characteristics by cohort

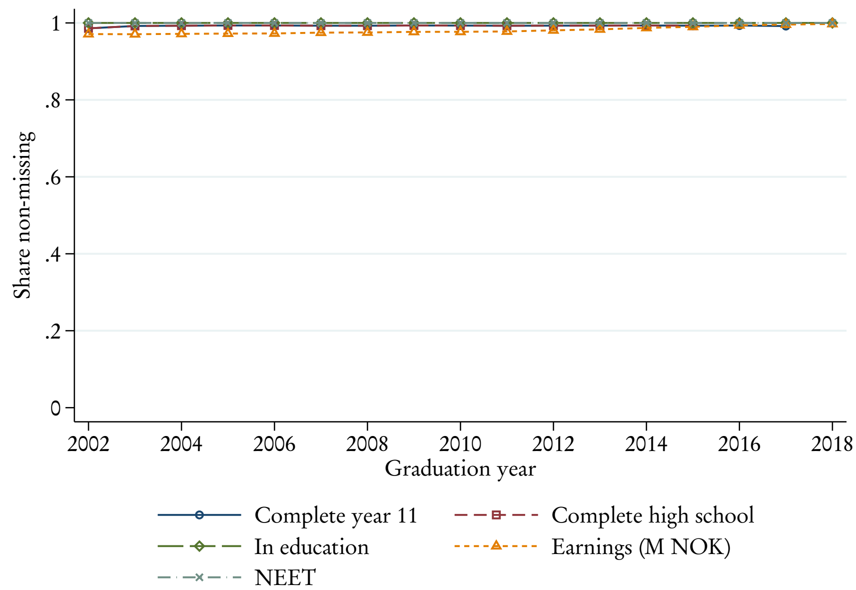


Figure A4: Share of students with non-missing longer-term outcomes by cohort

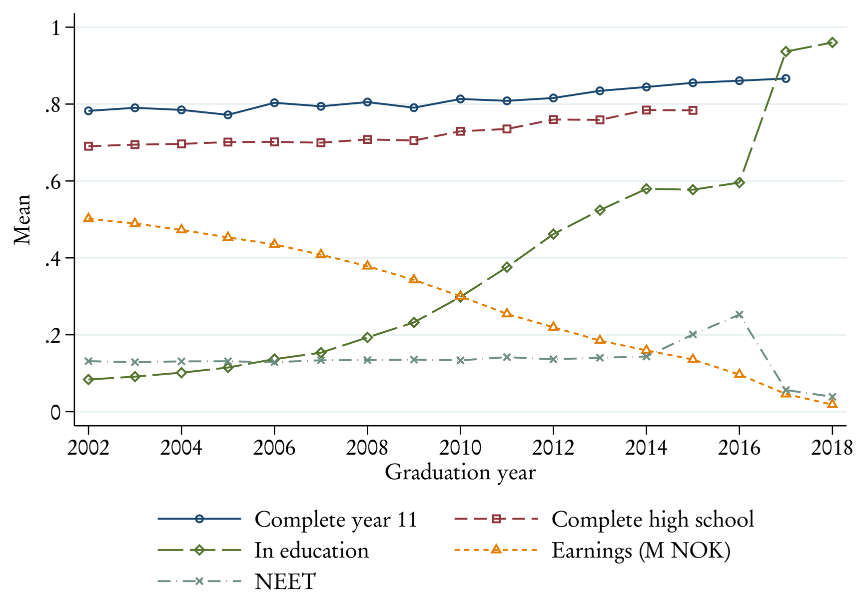


Figure A5: Mean longer-term outcomes by cohort

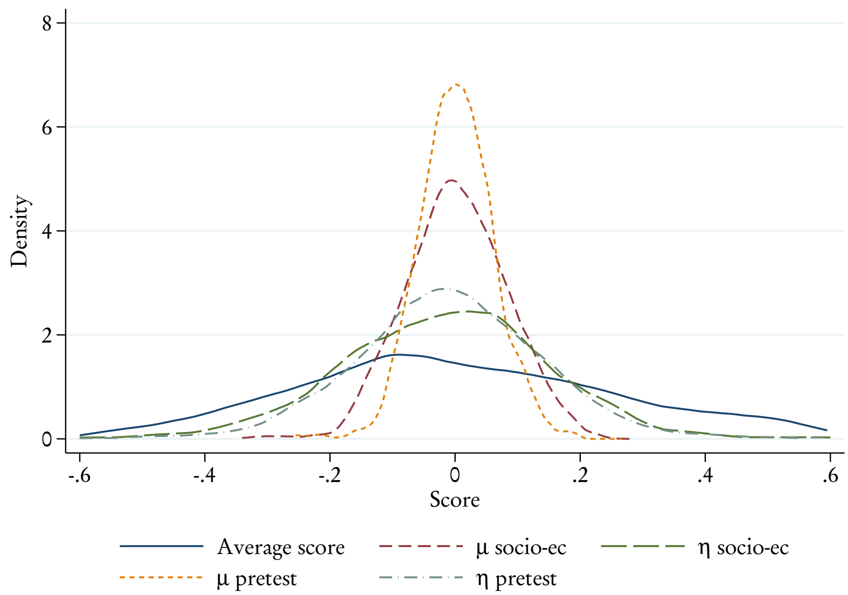


Figure A7: School-by-year average exam score and estimated VA for the 2015-2019 cohorts.

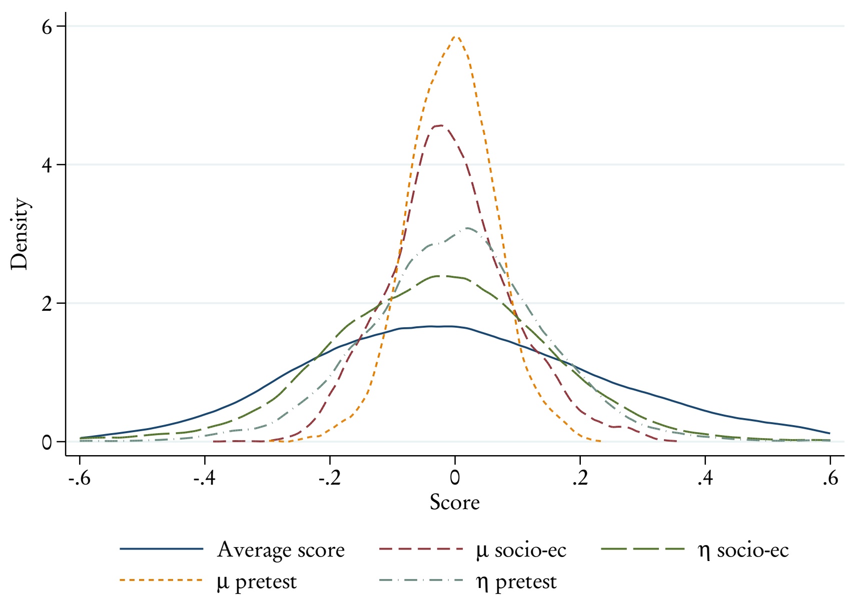


Figure A6: School-by-year average exam score and estimated VA for the 2004-2019 cohorts

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 upper secondary school	(5) Years schooling	(6) In education	(7) Employed	(8) NEET	(9) Earnings (NOK 100k)	(10) Log earnings
$\hat{\mu}$ written exam	0.990**	0.903**	0.904**	0.283**	1.263**	-0.002	0.096**	-0.073**	1.586**	0.214**
	(0.094)	(0.093)	(0.134)	(0.053)	(0.327)	(0.019)	(0.030)	(0.026)	(0.336)	(0.036)
$\hat{\mu}$ oral exam	0.920**	1.169**	1.254**	0.328**	1.188**	-0.006	0.073**	-0.068**	1.660**	0.208**
	(0.141)	(0.137)	(0.173)	(0.087)	(0.253)	(0.014)	(0.030)	(0.029)	(0.567)	(0.061)
$\hat{\mu}$ teacher	0.555**	0.697**	1.095**	0.203**	0.888**	0.004	0.065**	-0.064**	1.011**	0.109**
	(0.109)	(0.122)	(0.088)	(0.065)	(0.235)	(0.011)	(0.022)	(0.021)	(0.367)	(0.039)
Written exam score	1.000	0.514**	0.645**	0.151**	0.938**	0.012**	0.035**	-0.035**	0.643**	0.087**
	(.)	(0.004)	(0.005)	(0.002)	(0.018)	(0.001)	(0.002)	(0.002)	(0.018)	(0.003)
Oral exam score	0.491**	1.000	0.605**	0.150**	0.881**	0.012**	0.037**	-0.036**	0.631**	0.083**
	(0.004)	(.)	(0.005)	(0.002)	(0.018)	(0.001)	(0.002)	(0.002)	(0.015)	(0.003)
Teacher grade	0.716**	0.701**	1.000	0.238**	1.327**	0.013**	0.060**	-0.057**	0.868**	0.110**
	(0.005)	(0.003)	(.)	(0.002)	(0.018)	(0.001)	(0.002)	(0.002)	(0.028)	(0.004)
N	83372	83372	83145	82723	81154	83372	83372	83372	81135	72504
# clusters	428	428	428	428	428	428	428	428	428	428
$\bar{y}$	0.015	0.015	0.037	0.714	13.909	0.089	0.870	0.114	5.069	1.642

Note: See note to Table 2 Significant at * 10%, ** 5%

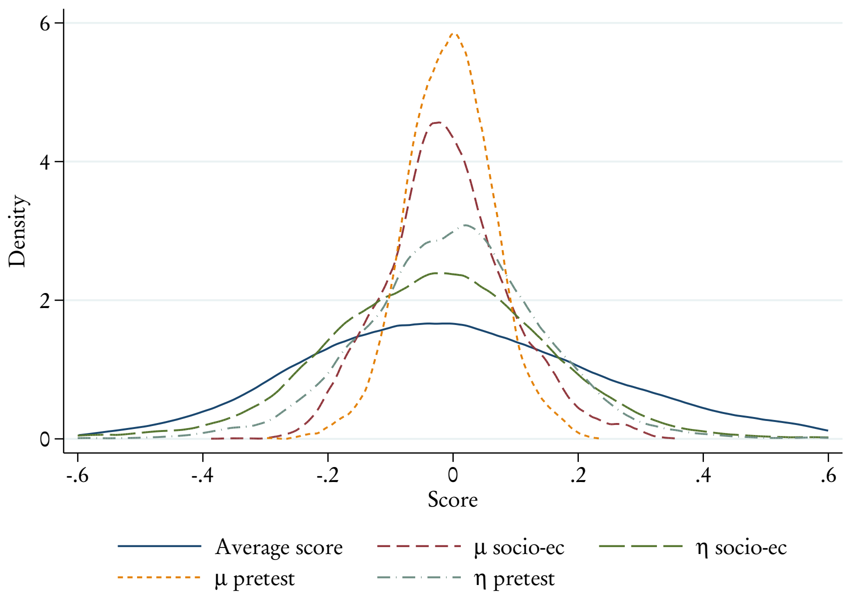


Figure A8: School-by-year average grade eight test score and estimated VA for the 2015-2019 cohorts

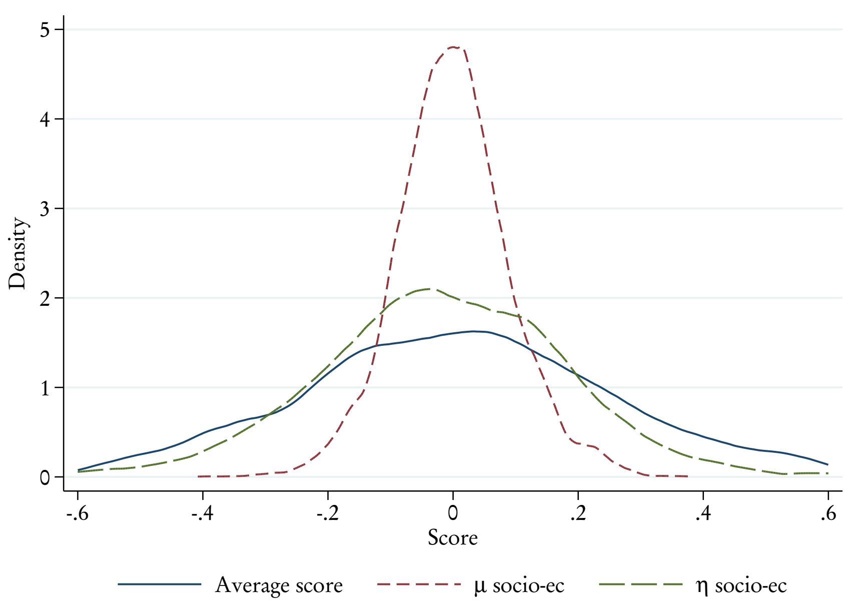


Figure A9: School-by-year average year five test score and estimated VA for the 2015-2019 cohorts

Table A2: Exam score VA and outcomes for school-changers, school fixed effects

	(1) Written exam	(2) Teacher grade	(3) Compl. upper sec. school	(4) NEET	(5) Control (pretest/ index)
Indicators controlling for pretest					
$\hat{\mu}$ year 8-school	-0.752	-0.633	1.778	-3.398	2.540
	(2.290)	(2.724)	(1.956)	(2.961)	(2.143)
$\hat{\mu}$ year 9-school	0.816**	0.457**	0.192	-0.034	-0.194
	(0.143)	(0.151)	(0.121)	(0.126)	(0.153)
$\hat{\eta}$ year 8-school	0.066	0.017	0.200	-0.290	0.243
	(0.192)	(0.229)	(0.163)	(0.248)	(0.174)
$\hat{\eta}$ year 9-school	0.290**	-0.018	-0.012	-0.019	-0.006
	(0.052)	(0.047)	(0.041)	(0.042)	(0.052)
Indicators controlling for family background					
$\hat{\mu}$ year 8-school	1.165	0.686	2.214	-2.044	1.174
	(1.725)	(1.988)	(1.484)	(2.014)	(1.177)
$\hat{\mu}$ year 9-school	0.723**	0.462**	0.115	-0.129	0.004
	(0.105)	(0.115)	(0.081)	(0.081)	(0.065)
$\hat{\eta}$ year 8-school	0.223	0.124	0.265*	-0.210	0.098
	(0.171)	(0.196)	(0.147)	(0.190)	(0.113)
$\hat{\eta}$ year 9-school	0.302**	0.074	0.017	-0.006	0.004
	(0.050)	(0.052)	(0.039)	(0.042)	(0.027)
Student controls	Yes	Yes	Yes	Yes	*
N students	8013	7827	4236	3229	8013
N clusters	935	932	842	798	935

Note: See notes to Table 6. Significant at * 10%, ** 5%

Table A3: Teacher grade VA and outcomes for school-changers

	(1)	(2)	(3)	(4)	(5)
	Written exam	Teacher grade	Completed upper secondary school	NEET	Control (pretest/ index)
Indicators controlling for pretest					
$\hat{\mu}$ year 8-school	0.014	0.058	0.029	0.026	0.058
	(0.088)	(0.081)	(0.070)	(0.070)	(0.084)
$\hat{\mu}$ year 9-school	-0.027	0.937**	0.059	0.010	-0.001
	(0.084)	(0.088)	(0.069)	(0.071)	(0.084)
$\hat{\eta}$ year 8-school	0.132**	0.159**	0.053	0.018	-0.028
	(0.053)	(0.054)	(0.040)	(0.039)	(0.048)
$\hat{\eta}$ year 9-school	-0.042	0.470**	0.043	-0.015	0.094*
	(0.051)	(0.051)	(0.041)	(0.037)	(0.051)
Indicators controlling for family background					
$\hat{\mu}$ year 8-school	0.150*	0.188**	0.079	-0.002	0.064
	(0.083)	(0.085)	(0.063)	(0.062)	(0.047)
$\hat{\mu}$ year 9-school	0.297**	0.980**	0.114*	-0.081	0.028
	(0.081)	(0.084)	(0.062)	(0.059)	(0.043)
$\hat{\eta}$ year 8-school	0.061	0.118**	0.051	-0.013	0.034
	(0.049)	(0.053)	(0.036)	(0.033)	(0.028)
$\hat{\eta}$ year 9-school	0.046	0.474**	0.078**	-0.033	0.018
	(0.048)	(0.050)	(0.037)	(0.033)	(0.025)
Student controls	Yes	Yes	Yes	Yes	*
N students	7954	7798	4198	3199	7954
N clusters	924	924	831	788	924

Note: 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 upper secondary school	(4) NEET	(5) Earnings	(6) Log earnings	(7) Background index
$\hat{\mu}^{Old}$	0.166**	0.325**	0.105**	-0.069**	0.670**	0.087**	0.121**
	(0.028)	(0.030)	(0.016)	(0.013)	(0.134)	(0.026)	(0.018)
$\hat{\mu}^{New}$	0.401**	0.721**	0.097**	-0.031**	0.277**	0.018	-0.047**
	(0.028)	(0.028)	(0.015)	(0.013)	(0.128)	(0.025)	(0.014)
$\hat{\eta}^{Old}$	0.086**	0.073**	0.035**	-0.018**	0.076	0.001	0.031**
	(0.014)	(0.015)	(0.008)	(0.007)	(0.065)	(0.012)	(0.007)
$\hat{\eta}^{New}$	0.390**	0.072**	0.058**	-0.007	0.213**	0.033**	0.003
	(0.014)	(0.015)	(0.008)	(0.007)	(0.063)	(0.012)	(0.007)
With neighborhood fixed effects:							
$\hat{\mu}^{Old}$	0.038	0.024	-0.005	-0.031	0.297	-0.074	0.035
	(0.064)	(0.066)	(0.037)	(0.034)	(0.432)	(0.083)	(0.030)
$\hat{\mu}^{New}$	0.342**	0.636**	0.071**	-0.024*	0.018	-0.032	-0.053**
	(0.031)	(0.031)	(0.017)	(0.014)	(0.157)	(0.030)	(0.015)
$\hat{\eta}^{Old}$	0.036**	0.029*	0.002	-0.012	0.024	-0.034**	0.025**
	(0.017)	(0.017)	(0.010)	(0.008)	(0.088)	(0.017)	(0.008)
$\hat{\eta}^{New}$	0.386**	0.066**	0.058**	-0.005	0.159**	0.028*	0.008
	(0.015)	(0.015)	(0.009)	(0.007)	(0.072)	(0.015)	(0.007)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	No
N students	95806	98228	82985	78055	38498	31200	104656
N clusters	10441	10477	10151	10028	8182	7617	10587

Note: See notes to Table 7. Significant at * 10%, ** 5%

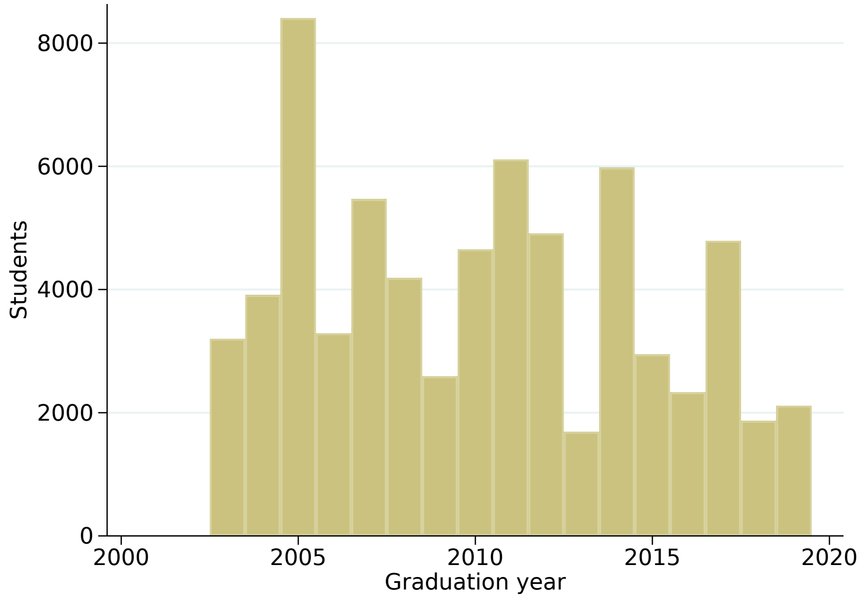


Figure A10: Year of change predicted school

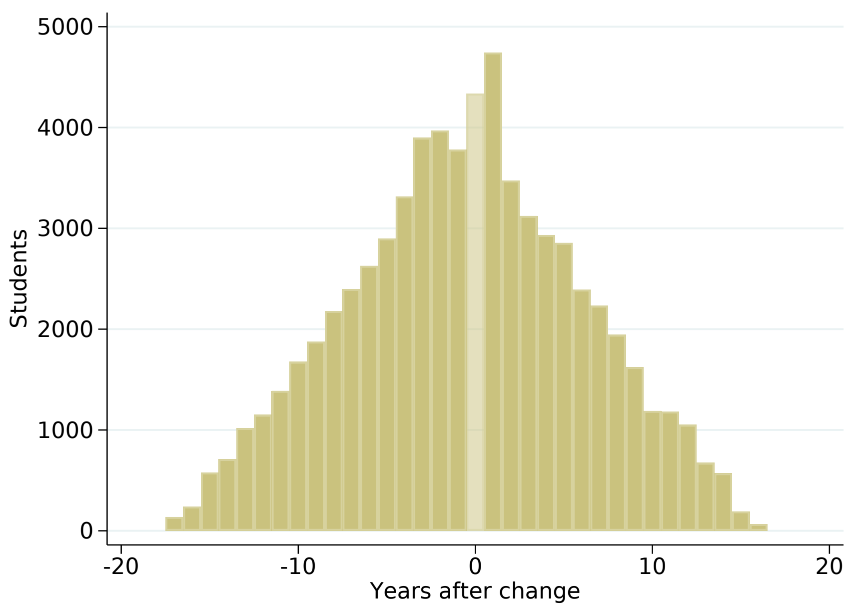


Figure A11: Years since change in predicted school

Table A5: Effect of change in assigned school quality on exam scores, neighborhood fixed effects

	(1) Exam score	(2) Teacher grade	(3) Complete upper secondary school	(4) NEET	(5) Background index
Students completing at new school					
$\hat{\mu}^{New}$	1.046**	0.568**	0.315**	-0.199**	-0.053
	(0.301)	(0.284)	(0.156)	(0.097)	(0.135)
$\hat{\mu}^{Old}$	0.155	-0.201	-0.131	0.008	0.302**
	(0.296)	(0.310)	(0.140)	(0.108)	(0.143)
$\hat{\eta}^{New}$	0.160**	-0.004	0.021*	0.000	0.012
	(0.024)	(0.022)	(0.012)	(0.008)	(0.009)
$\hat{\eta}^{Old}$	0.032	0.049	-0.011	-0.003	0.022
	(0.057)	(0.056)	(0.025)	(0.015)	(0.023)
Students completing compulsory schooling at old school					
$\hat{\mu}^{New}$	0.298	0.270	0.308*	-0.298**	-0.097
	(0.330)	(0.305)	(0.176)	(0.109)	(0.147)
$\hat{\mu}^{Old}$	0.658**	-0.211	-0.028	0.074	0.148
	(0.283)	(0.301)	(0.129)	(0.099)	(0.136)
$\hat{\eta}^{New}$	0.041	0.013	0.010	-0.027**	0.019
	(0.049)	(0.046)	(0.021)	(0.013)	(0.020)
$\hat{\eta}^{Old}$	0.251**	0.015	0.011	0.002	-0.020
	(0.027)	(0.026)	(0.012)	(0.008)	(0.013)
N students	66789	66160	54617	51633	66789
Nclusters	1212	1212	1206	1204	1212

Note: See notes to Table 8. Significant at * 10%, ** 5%

Table A6: Effect of change in assigned school quality (teacher grades) on exam scores

	(1) Exam score	(2) Teacher grade	(3) Complete upper secondary school	(4) NEET	(5) Background index
Students completing compulsory schooling at new school					
$\hat{\mu}^{New}$	0.160**	0.681**	0.024	-0.029	0.008
	(0.075)	(0.077)	(0.030)	(0.023)	(0.048)
$\hat{\mu}^{Old}$	0.010	0.089	-0.025	-0.018	0.033
	(0.063)	(0.076)	(0.031)	(0.023)	(0.048)
$\hat{\eta}^{New}$	-0.022	0.163**	0.014	-0.009	-0.061**
	(0.020)	(0.024)	(0.010)	(0.007)	(0.013)
$\hat{\eta}^{Old}$	0.043	0.080	-0.008	-0.016	-0.041
	(0.047)	(0.055)	(0.025)	(0.017)	(0.029)
Students completing compulsory schooling at old school					
$\hat{\mu}^{New}$	0.132*	0.092	0.063*	-0.012	0.057
	(0.079)	(0.072)	(0.036)	(0.020)	(0.044)
$\hat{\mu}^{Old}$	0.193**	0.847**	0.106**	-0.046**	0.025
	(0.069)	(0.063)	(0.033)	(0.018)	(0.041)
$\hat{\eta}^{New}$	0.037	0.048	0.016	-0.019	0.040*
	(0.044)	(0.041)	(0.021)	(0.014)	(0.023)
$\hat{\eta}^{Old}$	0.020	0.252**	0.034**	-0.003	-0.122**
	(0.021)	(0.021)	(0.010)	(0.006)	(0.011)
N students	66317	65990	54214	51243	66317
Nclusters	1205	1205	1199	1197	1205

Note: See notes to Table 8. Significant at * 10%, ** 5%