CESifo WORKING PAPERS

11216
2024

July 2024

# Alternative Measures of Teachers' Value Added and Impact on Short and Long-Term Outcomes: Evidence from Random Assignment

*Victor Lavy, Rigissa Megalokonomou*

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. 11216*

# Alternative Measures of Teachers' Value Added and Impact on Short and Long-Term Outcomes: Evidence from Random Assignment

## Abstract

A recent critique of using teachers' test score value-added (TVA) is that teacher quality is multifaceted; some teachers are effective in raising test scores, others are effective in improving long-term outcomes This paper exploits an institutional setting where high school teachers are randomly assigned to classes to compute multiple long-run TVA measures based on university schooling outcomes and high school behavior. We find substantial correlations between test scores and long-run TVA but zero correlations between these two TVA measures and behavior TVA. We find that short-term test-score TVA and long-run TVA are highly correlated and equally good predictors of long-term outcomes.

JEL-Codes: J240, J210, J160, I240.

Keywords: teacher quality, quasi-experimental random assignment, university quality, choice of university study, panel information on teachers, teacher value added.

*Victor Lavy*
*Department of Economics*
*Hebrew University of Jerusalem / Israel*
*victor.lavy@mail.huji.ac.il*

*Rigissa Megalokonomou*
*Department of Economics*
*Monash University, Melbourne / Australia*
*rigissa.megalokonomou@monash.edu*

We thank the AER editor, John Friedman, for his detailed and useful comments on the first submitted draft, which led to a substantial revision of this paper. We also thank Anjali Adukia, Bo Jackson, Andrew Bacher-Hicks, David Card, Aimee Chin, Adeline Delavande, Alexandra de Gendre, Josh Goodman, Mike Gilraine, Caroline Hoxby, Kevin Lang, Dick Murnane, Cristian Pop-Eleches, and Chrisovalantis Vasilakis. We also thank seminar participants at the 2022 Australasia Meeting of the Econometric Society, the 13th International Workshop on Applied Economics of Education, the 2020 NBER Summer Institute, the 2020 CESifo Conference on the Economics of Education, the 2021 Organisational Economics Workshop in Australia, the 2021 Labour Econometrics Workshop at Monash University, the 1st Moscow International Workshop on Applied Research in Labor Economics and Human Capital, the 2023 Western Economics Association conference at the University of Melbourne, participants at the National and Kapodistrian University of Athens, University of Queensland, Queensland University of Technology, University of Essex, University of Adelaide, University of Wollongong, Bangor University, E61 Institute, University of Warwick, and Hebrew University. We are grateful to Elias Arvanitakis for helping compile the data, and to Dominic Byrne and Yuancheng Han for excellent research assistance. This paper uses data collected from school archives and administrative data and we are not at liberty to publish online; however, we can provide the code as well as detailed instructions and assistance on how to apply and access the data to enhance the replicability of our analysis. There is no material or financial conflict of interest on the part of any of the authors with respect to any of the research in this paper. Rigissa Megalokonomou acknowledges research support from the University of Queensland BEL Early Career Grant (No: UQECR1833757) and the Monash University, Monash Establishment Grant (No: 1755774).

# 1 Introduction

Teachers' value added (TVA) has been shown to affect students' outcomes (Chetty, Friedman, and Rockoff, 2014a,b). This methodology decomposes students' test scores into components attributed to student heterogeneity and teacher quality. One concern is that TVA is measured based on students' test scores' improvements in the short term. Teachers' contemporaneous influence on students' short-run outcomes, such as test scores, is not necessarily a good measure of the lasting impact teachers may have on students' longer-term outcomes. Teachers may affect students' long-term outcomes that do not run through short-term tests. For example, longer-term success may require a more substantial change in aspirations and motivation beyond performing well on an exam. Thus, teacher quality might be multifaceted; some teachers are effective in raising test scores, and others are effective in improving long-term academic or behavioral outcomes. We follow recent developments in the literature and derive TVA measures of teachers based on long-term and behavioral outcomes (Gilraine and Nolan, 2021; Rose, Schellenberg, and Shem-Tov, 2022; Petek and Pope, 2023). These capture teachers' effectiveness in improving high school students' novel long-term (university admissions) and noncognitive (behavioral) outcomes. Using these multiple TVA measures, we assess whether teachers may have lasting effects on long-term outcomes beyond the effect on short-term outcomes, such as test scores. These measures also enable us to explore potential correlations between the various TVA metrics.

Significantly, we advance the estimation of multiple TVA measures in a context where teachers are randomly assigned to classes. In some experimental settings, TVA's estimated effects are unbiased (Kane, McCaffrey, Miller, and Staiger, 2013; Kane and Staiger, 2008). However, several studies have raised doubt about whether measured TVA and its estimated effects are also unbiased when using observational data (Rothstein, 2009, 2010; Koedel and Betts, 2011; Kinsler, 2012; Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson, and Shepard, 2010; Rothstein, 2017). Strong assumptions about the nature of the educational production function and students' classroom assignment are implicit in the value added approach. In particular, the explicit assumption in these models —that teachers are randomly assigned to students—does not hold in many settings. Therefore, TVA estimates may capture more than teachers' effectiveness in improving the respective students' outcomes.[1]

Several papers document non-random student-teacher matching and show that sorting based on teachers' quality and students' unobserved potential outcomes (initiated by parents or administrators) may be relevant in the assignment of teachers to classes (Bacher-Hicks, Chin, Kane, and Staiger, 2017; Horvath, 2015; Paufler and Amrein-Beardsley, 2014; Jackson, 2014; Sass, Hannaway, Xu, Figlio, and Feng, 2012; Koedel and

[1]These studies use an experimental design to show that TVA accurately predicts student achievement. However, there are concerns about externalizing this finding due to compliance issues and the random-assignment experiments were performed using pairs of teachers whose principals agreed to assign students to their classrooms.

Betts, 2011; Rothstein, 2010, 2009; Aaronson, Barrow, and Sander, 2007; Clotfelter, Ladd, and Vigdor, 2006; McCaffrey, Lockwood, Koretz, Louis, and Hamilton, 2004; Rothstein, 2017). In many contexts, these assignments are complex; they may be affected by parental requests and influence and students' needs and social dynamics should be taken into account. It also depends on teachers' comparative advantages concerning students' abilities and needs.

Another major issue that we can address in our setting is the reliability of TVA estimates, since they might lack stability if a large share of the value added measure is attributable to the student or classroom-specific factors unrelated to teachers (Lockwood, McCaffrey, Hamilton, Stecher, Le, and Martinez, 2007; Schochet and Chiang, 2013). Even if the measures are unbiased, there are concerns about whether they are stable enough from one year to another (Green III, Baker, and Oluwole, 2012; Darling-Hammond, 2015; Newton, Darling-Hammond, Haertel, and Thomas, 2010; Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson, and Shepard, 2010; Murnane, Willett, Duhaldeborde, and Tyler, 2000; Blazar, Litke, and Barmore, 2016; Stacy, Guarino, and Wooldridge, 2018).

Lastly, our empirical context allows us to address the critique that not many studies measure TVA impacts in high-stakes settings, which may be very different from low-stakes ones, in which students and/or teachers may not exert much effort in studying or teaching (Corcoran, Jennings, and Beveridge, 2011; Hanushek and Rivkin, 2010). These issues may be more problematic when value added models are used for evaluation. These concerns underline the prevailing view that value added models help classify and evaluate teachers but may lack causal interpretation (Koedel, Mihaly, and Rockoff, 2015).

In summary, several unique features of our setting enable us to examine in this paper the impact of teachers' quality on students' short-term outcomes (test scores in high school) and longer-term outcomes (university schooling and choice of major), escaping limitations faced in the related studies. In this context, teachers and students in public education are randomly assigned to different classes. This policy has been in place in Greece for many years, includes all public high schools, and results in random matching between teachers and students. This allows estimation of the causal effect of TVA on students' outcomes free of potential bias due to teachers' and students' endogenous sorting to classes. We use this population-based "experiment" to examine whether this relationship varies by schools, teachers, and students' characteristics, such as gender. We first use short-term outcomes to measure TVA and then examine the impact of those TVA measures on students' outcomes. We then use novel long-term outcomes to measure teacher effectiveness. The first is the threshold score used for admission to the student's study program. The second is the quality rank of the enrolled university degree. Based on the potential multidimensional teacher's quality, we also estimate a TVA that captures student disruptive behavior in the class to examine the behavioral impact of teachers on students (Petek and Pope, 2023). We find substantial correlations between short-run and long-run value added measures, but small correlations between test-score value added and behavior value added measures. In the paper's last section, we examine the nature of the relationship between TVA and teachers' entry and exit. This analysis sheds light on whether schools use TVA in decisions regarding teacher retention.

We use data from 21 schools that form a relatively representative sample of high schools in Greece. The data cover the period 2003-2011. These data include test score information for the $10^{th}$-$12^{th}$ grades on national and school exams. The sample includes 449 unique public school teachers who teach along the whole spectrum of the high school curriculum. We combine these data with information from the Ministry of Education on enrollment in universities in Greece. These data include enrollment, institution, and field of study in universities and other higher education institutions. We also obtained information from the State Scholarships Foundation about whether students won a State Government Scholarship for their undergraduate studies.

Our setup in this study has several advantages that circumvent the problems and difficulties other studies encounter when using observational data to measure TVA and its effects. First, the random assignment of teachers and students prevents sorting between students and teachers. Second, we observe multiple teachers who are teaching the same grade in the same school in a given year.[2] This enables us to account for school-level unobserved shocks that vary at year or grade level. Third, we observe several teachers who are teaching each student in the same grade. This means that we have test scores and teacher-student matches in many courses (subjects) and multiple grades ($11^{th}$-$12^{th}$). This data structure allows us to control for year, school, grade, and class variation by fixed effects when we estimate TVA. When estimating TVA's effect on student outcomes, we can also use a student fixed effects specification and its interaction with the other fixed effects included in the model. Fourth, we track students over time and follow them through their transition to tertiary education. Thus, we have information on their university admission outcomes and know which teachers they had in each grade and subject. This allows us to measure the quality of the same teacher using multiple value added measures.

Our baseline results involve test-score value added measures. We examine the correlation between these TVA measures and TVA measured based on non-test score outcomes. We find that students assigned to high-TVA teachers in high school have higher test scores on national exams. Since these test scores are used in admissions to higher education, a consequence is that these students are also more likely to continue postsecondary schooling in universities than in vocational or technical institutions. They are also more likely to be admitted to higher-ranked universities and more competitive study programs. Remarkably, higher-TVA

[2] Few studies have combined the measurement of TVA with systematic observations for each teacher (Kane, Staiger, and Seattle, 2012).

teachers in high school also affect the student's choice of field of study at university. The effect size of these impacts is meaningful economically. For example, a 1 SD improvement in TVA raises normalized average test scores by approximately 0.20 SD in $11^{th}$ and $12^{th}$ grades. It increases the likelihood of continuing university education by 7 percentage points. Also, a 1 SD improvement in TVA renders students 1 percentage point more likely to obtain a state government scholarship for their undergraduate studies. These effects do not vary by student gender and are robust to various conditioning variables. We also find that the impact of TVA on test scores is increasing in the TVA quintile, with teachers in the top quintile of TVA having a significantly larger impact than teachers in the bottom quintile of TVA. We also apply two correction methods to our main estimates on subject-specific national exams as a robustness test: an empirical Bayes (EB) shrinkage approach and a two-step bootstrapping technique. We do so to address potential sampling error and account for the variable of interest (TVA measure) being a generated regressor. Our results are relatively unchanged when using these two alternative strategies.

We then measure the value added by teachers using student longer-term outcomes by replacing the (residualized) test scores with different university admissions-related outcomes from 1 year after a student was in a teacher's class. We find substantial correlations between test scores and long-run TVA measures, varying from 0.6 to 0.9 depending on the long-run outcome. We also find that being randomly assigned to teachers with high long-run TVA positively impacts student outcomes. In particular, we find that long-run value added measures positively impact students' long-run outcomes. We then measure value added based on teachers' ability to reduce students' suspensions one year after being assigned to their class. Recent studies show that noncognitive/behavioral value added measures may be good predictors of specific types of outcomes: long-term behavioral outcomes (Jackson, 2018; Rose, Schellenberg, and Shem-Tov, 2022; Petek and Pope, 2023). We test whether behavioral value added measures affect academic outcomes. However, our behavioral value added is only based on suspensions, while other studies include several behavioral and GPA variables (Jackson, 2018; Petek and Pope, 2023; Gilraine and Nolan, 2021). We find that assignments to teachers who effectively reduce students' suspensions positively impact student outcomes. Also, we find zero correlation between the test-score and our noncognitive/behavioral value added measure. We then examine how well short-term test-score TVA, long-run TVA, and behavior TVA predict students' longer-term success while including all those TVA measures in the same regression. We find that all TVA measures predict future student outcomes. When we compare their relative magnitudes, we conclude that long-run TVA and short-term test-score TVA are the best predictors of future student success, although the impact of the long-run TVA is less precisely measured.

There are several potential mechanisms through which TVA might affect test scores. We present suggestive evidence showing that these effects are partly driven by reducing unexcused absenteeism from school and thus effectively increasing hours of instruction. On the other hand, we find no effect on excused absenteeism due to illness or other reasons that are not self-chosen.

In the last part of the paper, we examine the relationship between TVA and the entry and exit of teachers from schools. We also examine whether this relationship varies by teachers' gender and school quality. We obtain suggestive evidence that high-TVA teachers are less likely to be retained in schools, and this average effect is due to the mobility dynamics of teachers in low-performing schools. Males and new teachers appear to be more mobile. These findings are potentially important in discussing which schools can keep their high-productivity teachers and attract new ones with high potential value added.

This paper contributes to the literature on teachers and schooling quality in several important dimensions. First, our study contributes to the recent literature on using long-run value and behavioral value added measures to capture teacher effectiveness and examine correlations with standard test-score value added (Gilraine and Nolan, 2021; Petek and Pope, 2023; Rose, Schellenberg, and Shem-Tov, 2022). This is paramount, since teachers seem twice as important for improving longer-term postsecondary-related outcomes than previously estimated. Our study is the first in which students' test scores are measured in many subjects and on high-stakes exams that matter for other adult outcomes. For example, we present evidence on the impact of teachers' effectiveness in STEM versus non-STEM fields of study—an important distinction, given the growing importance of STEM education for individuals' labor market outcomes and nations' economic growth. Our study is the first to show high persistency in same-teacher VA measures across time, in different classes, and in grades. Also, the more balanced gender proportion of teachers in the Greek education system presents a unique opportunity to distinguish between female and male teachers' productivity. We find that the productivity of teachers in a high-stakes environment does not vary by gender. We are unaware of another study that considers such evidence, which is likely important for understanding disparities in gender earnings and career advancement.[3] Another unique aspect of this paper is that we measure teachers' impact beyond test scores, particularly on the choice of field of study, and find that this effect is the same for male and female students. This contributes to the growing literature on the choice of field of study. Studies on this important topic have focused mostly on the expected returns in the labor market, with less attention paid to behavioral factors, especially in the high school environment. The last item in this list of 'contributions' is identifying TVA's effect on students' outcomes without any potential confounding effect due to sorting or selecting teachers or students. The random placement of teachers and students in classes yields a random experimental construct. The study by Kane, McCaffrey, Miller, and Staiger (2013) is among the very few we

[3] This result is in line with findings reported by Lavy (2013), who studied gender differences in competitiveness among teachers and found no such disparities.

know that randomly assigns students to teachers in school and estimates the effect of teachers' effectiveness.[4]

## 2 Context

### 2.1 High Schools in Greece

The Greek education system is administered by the Ministry of Education, Research, and Religious Affairs and is similar to other southern European systems. It is highly centralized and directly managed by the Ministry of Education, which controls curricula, teaching materials, and examinations. There is a very small private sector. Students are assigned to public schools through zoning based on their residential address and residential proximity to the school. The study curriculum in high schools in Greece includes core courses (modern Greek, mathematics, physics, biology, and history) and courses in one of three tracks (classics, science, or exact science) in all high schools. Students choose a track at the beginning of $11^{th}$ grade and can change it in $12^{th}$ grade, although very few do. No performance threshold exists for a student to enter a high school track. All schools that administer national exams follow the same curriculum and offer courses in core and track subjects based on the material covered on the national exams (OECD, 2018).

National exams include core and track subjects. Until 2005, students took national exams in $11^{th}$ and $12^{th}$ grades. Since 2006, the national exams have been administered only at the end of grade 12. The Ministry of Education receives national exam answer sheets and sends them to examiners nationwide with the student's information concealed. Schools also administer similar-format school exams that cover the same curricula as the national exams, but the students' teacher grades them. Most questions on the national and school exams are open, and only a few can be multiplechoice. For 2003-2005, we observe students' school and national scores in both grades ($11^{th}$ and $12^{th}$). From 2006 to 2011, the data include only $12^{th}$-grade school and national exam test scores. Schools use school and national exam test scores to determine grade completion and high school graduation. These test scores also feature in decisions regarding grade repetition and appear on the high school graduation diploma, which employers often use in hiring decisions.

### 2.2 Admission to Postsecondary Education

Postsecondary education is free in Greece. There are university entrance exams, but no tuition is charged.[5] This is because the Greek constitution says that all Greeks (and some foreigners) are entitled to free education. Public schools and universities even provide free textbooks to all students. Most undergraduate degrees in Greek postsecondary institutions take 4 years to complete on time. Students enroll in academic universities

[4]An exception is Feld, Salamanca, and Zölitz (2020), who use data from a Dutch business school to calculate value added measures of tutorial instructors randomly matched with university students conditional on scheduling conflicts.

[5]Students who fail university admissions exams can retake these exams 1 year later (Bizopoulou, Megalokonomou, and Simion, 2022).

or vocational schooling.

### 2.3 Teachers

The Ministry of Education in Greece exercises control over state schools regarding curricula, staffing, and funding. Teachers are civil servants and receive a salary based on seniority, location, and family size. A teacher's salary structure is similar to that of other employees in the public sector and has been a point of contention with teachers' unions (Stylianidou, Bagakis, and Stamovlasis, 2004). There are two tracks for teachers: permanent and temporary or substitute teachers.[6] Teaching needs are first met by using existing permanent staff, hiring temporary staff, and hiring a permanent teacher as a last resort. Temporary teachers get paid on the same scale as entry-level permanent teachers. Permanent staff is difficult to fire, especially in public schools. Teachers can be fired for their inability to do their job, but documenting this is difficult (**?**).

Teachers who want to be moved to other schools submit a request to the Ministry of Education and are assigned to the waiting list. Schools with excess teaching needs must submit a request to the Ministry of Education, which is responsible for hiring and assigning teachers to schools based on subject-specific teaching needs.

### 2.4 Random Assignment of Teachers and Students to Classes

Students and teachers are assigned to classes within each school in a way that results in students' and teachers' random matches. Once students enroll in a given high school, they are assigned to a physical classroom where they take all core courses.[7] Students' assignments to classrooms are based on their surnames' alphabetical order.[8] Assignments based on ability, family background, or any other observed characteristic are strictly prohibited. The school principal implements the lexicographic assignment of students to classes, which is maintained throughout all high school grades. Law Number 1566 states that schools should be the focal point of integration for students of different backgrounds, genders, and abilities. The same law states that the school should contribute to the "holistic, harmonious and balanced development of the pupils' mental and psychosomatic attributes." The aim is that all students—independent of gender, ethnicity, and ability—should evolve into complete personalities and develop their skills in a social environment that does not separate students based on specific characteristics. Thus, their class assignment must be based solely on

[6] Permanent teachers are considered to be civil servants and enjoy job security. Temporary or substitute teachers are contracted for up to 10 months and have to reapply via a centralized assignment system for a new short-term appointment (Dinerstein, Megalokonomou, and Yannelis, 2022).

[7] Students are assigned to different classes for optional and track subjects.

[8] Presidential Decree 323 and Γ2/47380/ states that students must be allocated to classes based on a strict alphabetical order not only for their daily classes, but also for all examinations in the school.

alphabetical surname order. Students are not allowed to switch to another class based on preferences.[9]

Presidential Decree 201 states that the school board must annually assign the next academic year's teachers to classes in June. Several rules guide this process. First, it should facilitate teachers' teaching schedules, considering their subject specialization. Second, the school should avoid a teacher's assignment to the same class in two consecutive grades. Third, teachers can teach the same class twice during the three high school grades.[10] Fourth, this assignment should be unrelated to teacher status (permanent/ temporary/ hourly) and teaching experience (years in the profession).[11] The law states that if there is any disagreement within the school board about teachers' assignment to classes, a representative of the school authority and school counselor have the final word in this decision.[12] We present supporting evidence below to confirm that these rules result in a quasi-random match between students and teachers.

### 2.5 The Greek University Admissions System

The education system in Greece is highly centralized (OECD, 2018). The Ministry of Education administers university admissions and most Greek universities are public. Most undergraduate degrees in Greek universities can be completed in four years, except the Polytechnic University in Athens, which takes five years.[13] Student performance on the exams described in Section 2.1 are used to determine postsecondary admissions.[14]

University applicants submit a list of their preferred postsecondary programs[15] to the Ministry of Education (OECD, 2018).[16] The admissions cutoff score for each study program is unknown to students when applying. Students can apply to several programs at several universities, and application to various programs is not restricted by high school track.[17] However, postsecondary programs assign different weights to different high school subjects when computing applicants' university admissions scores.

[9]In experimental studies where students were randomly assigned to teachers with different TVA measures, an issue of low compliance with randomization blocks among participating students was observed. In particular, some students (or parents) switched out of their randomly assigned teacher's classroom to a classroom in which the teacher was nonrandomly selected. An instrumental variable approach was used in those cases to estimate the treatment effect on the treated (Kane, McCaffrey, Miller, and Staiger, 2013; Bacher-Hicks, Chin, Kane, and Staiger, 2017). Our design is not sensitive to noncompliance issues, since students cannot move between classes.

[10]Only under very special circumstances would teachers be allowed to teach the same class three times (consecutively or not) and this must be approved by the school board and the school counselor.

[11]Also, teachers cannot be assigned to a class in which their child or any other relative is a student.

[12]School counselors are associated with the Ministry of Education and work closely with the teaching staff, school principal, school council, director of the local school authority, parents' association, and government.

[13]This is the most prestigious engineering program in Greece.

[14]The exams' scores are weighted unevenly: 70% for the national exam and 30% for the school exam. Until 2005, university admissions were based similarly on $11^{th}$-grade school and national exams.

[15]By "program", we mean the university and field of study.

[16]See Goulas and Megalokonomou (2018) for more details on the admissions algorithm.

[17]For instance, students from any track can apply to an economics study program as long as they take the optional course in economics in $12^{th}$ grade. Admission to an engineering or science study program depends crucially on test scores in mathematics and physics in the science track.

## 3 Data

The data we use in this study come from schools and other administrative sources. We use data from a relatively representative sample of high schools in Greece.[18] The sample includes public schools from large and smaller urban and rural cities. These schools are located in different parts of Greece, as shown in Figure A1. The baseline sample is $11^{th}$-grade students in 2003-2005 and $12^{th}$-grade students in 2003-2011. Our primary analysis uses information on teachers, students, and principals from 21 high schools. The teachers' information permits us to track individuals' teaching history in a specific school from 2003 to 2011. This includes yearly class assignments (year, grade, class, and subject) from panel data on teachers. We infer teachers' gender from their first name, which forms a unique match in Greek.

### 3.1 Students

We obtained student-level information from each school, including a unique student identifier, gender, year of birth, study track, absenteeism records in grades 11 and 12, and test scores from school and national exams in all subjects in $11^{th}$ and $12^{th}$ grades. We also obtained $10^{th}$-grade school exam scores, which we use as a lagged test score in the TVA estimation. Test score data are available for 17 subjects. The raw school exam is also on a 1-20 scale, and we transform it into z-scores for each year, school, type of exam, and subject. We also obtained data on class size in each course.

Unique student identifiers allow us to match $10^{th}$ grade students to their $11^{th}$ and $12^{th}$ grade classes and test scores on the national exams. The dataset uses a panel structure for each student, which includes multiple observations by grade and subject. We then link the students' and teachers' datasets with administrative data from the Ministry of Education, which contains postsecondary schooling information. The latter includes the student's enrollment in any postsecondary schooling institution, the institution's name, the area studied, the enrolled department's quality rank, and the complete list of all departments/institutions to which the student applied.

Tables 1 and A2 present descriptive statistics for the sample used in the primary analysis in 2003-2005

[18] We show evidence of that in Lavy and Megalokonomou (2024) and also in Table A1—in which we compare the school characteristics in our sample with those of the remaining schools in the country and find no differences. The same sample of schools and teachers is also used by Lavy and Megalokonomou (2024) and Dinerstein, Megalokonomou, and Yannelis (2022). During 2011-2014 large effort was exerted to collect data on students in a single grade from a large sample of schools in Greece. The Ministry of Education and the corresponding Ethics Committee that evaluated the project permitted researchers to access a random sample of schools. The researchers were provided with a long list of school codes generated by a computerized algorithm and visited those schools. Data from many schools (representative sample) were obtained (more than 10% of the school population). More recently, those schools were contacted again and asked to give the researchers access to their teacher-level data and allow the matching of student to teacher data in their facilities. In several schools, the school management (principal, school authority, or school board director) had changed. As a result, they were unfamiliar with our research and refused to participate. We obtained the matched teacher-student data from a smaller number of schools, which we use in this paper.

($11^{th}$ and $12^{th}$ grades) and 2003-2011 ($12^{th}$ grade only). In 2003-2005, the proportion of female students was 56% and students were 17 years old, on average. Students study, on average, 16 subjects in $11^{th}$ and $12^{th}$ grades. Roughly 35% of students are in the classics track, 24% in the science track, and 42% in the exact science track in the $11^{th}$ and $12^{th}$ grades. The average class size is 21; 81% of students enroll in university schooling, 48% in academic universities, and the rest in technical education institutions. Table A2 shows that these figures were similar in 2003-2011. Students study nine subjects in $12^{th}$ grade on average.

### 3.2 Postsecondary Schooling Outcomes

The Ministry of Education collects information on each student's postsecondary applications and admissions. We have information on enrollment in postsecondary schooling (binary indicator=1 if enrolls) and enrollment in an academic or technical school (binary indicator=1 if academic university). Also, we use a measure of postsecondary schooling quality and a measure of how preferred the enrolled postsecondary institution is for the student. We derive each postsecondary program's annual admissions cutoff using the minimum score of the last-ranked enrolled student; this is the official program admissions cutoff or threshold the Ministry uses. We derive this variable using the 2003 university admissions cutoffs, the first year we have data on.[19] The most selective university degrees include studies in engineering and medicine, and the least selective include geo-technology and environmental studies. We also have available information on students' desired program preferences (field of study and institution) based on students' reported preferences over the university degrees they include in their preference list. The Ministry of Education uses a computerized central algorithm to rank all students based on their university admissions score and assign them to their most preferred program based on availability.

We also employ student-level scholarship data from the State Scholarship Foundation. Each year, the top 1% of students entering university in each degree receives a merit scholarship for the period of study. This is a selective and prestigious state scholarship that less than 2% of the student population in the country receives on average.

### 3.3 Teachers

We observe the teaching record for each teacher during the study period; this includes the class, grade, subject, and course in each of the years a teacher appears in our sample. Our focus is naturally on courses that lead to national $11^{th}$ and $12^{th}$ grade exams. Core subjects include modern Greek, history, biology, physics, and mathematics. Classics track subjects include Ancient Greek, Philosophy, and Latin in $11^{th}$ grade and ancient

[19] An alternative university degree quality measure comes from using the mean performance of students who enroll in each program each year. The two university degree measures have a simple correlation of more than 0.95. All results are similar if we use the alternative postsecondary program quality measure.

Greek, Latin, literature, and history in $12^{th}$ grade. Science track subjects are mathematics, physics, and chemistry in $11^{th}$ grade and biology, mathematics, physics, and chemistry in $12^{th}$ grade. Exact science track subjects are mathematics, physics, and computer science in $11^{th}$ grade and mathematics, physics, business administration, and computer science in $12^{th}$ grade. This information is available for 449 teachers in the sample of 21 high schools. The matching of students and teachers allows us to build a panel data set on teaching history for each teacher in the three grades. We use this data structure to construct a measure of teaching experience for each teacher during our study period. This measure is the number of unique combinations of classes-subjects-years-grades (workload) teachers taught during the sample period.

Table 1 presents descriptive statistics for teachers in our 2003-2005 sample: 51% are female, and the average experience based on previous teaching workload is 10 classes-subjects-years-grades combinations, but with considerable variation from 1 to 42. In the 2003-2011 sample, 50% of teachers were female, and female teachers' TVA was not much different from that of males (as shown in Table A2). Table 2 shows that each student has, on average, five different teachers in $11^{th}$ grade for whom we computed TVA and six such teachers in $12^{th}$ grade. Table A3 shows that students have six teachers with computed TVA in $12^{th}$ grade in the 2003-2011 sample.

### 3.4 Evidence on the Random Assignment of Students and Teachers to Classes

To further support our claim of students' and teachers' random assignment to classes, we test whether the teacher characteristics we observe are correlated with the available pre-assignment student characteristics (lagged test scores, age, gender). Teacher characteristics include gender, previous year's TVA (in the following section, we provide details on how we compute TVA), and teaching experience (measured by teacher's workload in the sample period).

We check the balancing implications of the random assignment of teachers and students in two ways. First, we regress the three pre-matched student characteristics and prior test scores (previous-year GPA, previous-year test scores in mathematics and English) on each of the four teacher characteristics: grade, year, track, and class fixed effects. Second, we regress each teacher's characteristics on all three student characteristics conditional on grade, year, track, and class fixed effects.[20] The first balancing exercise is shown in Table 3 for 2003-2005. Column (1) presents estimates from regressions of students' previous-year GPA on each characteristic of teachers. Columns (2), (3), (4), and (5) present the estimates from separate regressions of students' prior performance in mathematics or English, gender, and age on each teacher characteristic. All estimates are small (most are close to 0) and not significantly different from zero. These results suggest no

[20] Goulas, Griselda, and Megalokonomou (2022a,b); Goulas, Gunawardena, and Megalokonomou (2023); Goulas, Megalokonomou, and Zhang (2022a,b) also provide evidence that students are randomly assigned to classrooms within school cohorts in Greece using a larger sample of schools and students.

significant correlations between teacher and student characteristics.

Table A4 in the Online Appendix presents estimates for our second balancing exercise. Each column in the table presents estimates from regressions of each teacher's characteristics on all student characteristics, while all student characteristics are included in each regression. Outcome variables are reported horizontally. Almost all estimates presented in the four columns confirm that teacher characteristics are unrelated to student characteristics, since one estimate out of 20 is statistically significant at a 10% level and are all very close to 0. Overall, these results reassure us that teacher assignment is not systematically correlated with students' characteristics. As another check, we report the F-statistic (and the corresponding P-value) for the joint significance of the coefficients. We also show these estimates separately by track (in Tables A5 and A6) and core (in Tables A7 and A8) classes, as well as for the 2003-2011 period (in Tables A9 and A10). All of these tables provide additional evidence of the random assignment of students and teachers to classrooms.

A second and meaningful way to further investigate the balance between treatment status and potential outcomes is by estimating the correlations between TVA (details below) and student ability. One might worry that high-TVA teachers are systematically assigned to low-performing classes (or the opposite). In Figure 1, we present the TVA distribution by decile of students' previous-year's test scores in the same subject. There are 10 deciles of previous-year test scores, with 1 being the lowest decile and 10 the highest. We compute assigned teachers' TVA distribution for each decile of previous-year test scores. This distribution is relatively flat, with very small deviations from zero. The confidence intervals shown in the figure are all symmetrically drawn around zero. This additional evidence indicates that TVA and student ability are not correlated systematically.

The random assignment of teachers to classrooms relies on the assumption that there is more than one available teacher per teaching specialty to be assigned to a classroom within school-grade-cohorts. Also, there must be more than one classroom slot per teaching specialty to be filled. We provide suggestive evidence that there is enough variation in classroom slots that need to be filled by teachers and available teachers by teaching specialty per school, grade, and year in our data. In Table A11 we show summary statistics for classroom slots availability in Panel A and for teacher availability in Panel B. The three main teaching specialties cover most subjects taught in the related grades: (a) mathematics, which includes all courses in mathematics, geometry, and algebra in the core and the track; (b) physics, chemistry, and biology, which includes all courses in physics, chemistry, and biology in the core and the track; and (c) language and history, which includes all courses in modern Greek, ancient Greek, Latin, history, philosophy, in the core and the track. We find that there are around 10 classroom slots in each teaching specialty that need to be filled by teachers in each school, grade, and year, and 2-4 available teachers in each teaching specialty to be assigned to classroom slots in each school, grade, and year, on average. This reassures us that there is sufficient variation in the allocation process and that the random assignment is meaningful.

The degree to which TVA estimates are biased depends fundamentally on the extent to which students are sorted into teachers. The cumulative findings in Tables 3, A4–A10, and Figure 1 demonstrate that teachers and students are randomly matched in Greece. Therefore, we can view this setup as a "population-based" random experiment in which OLS regressions can be used to estimate the causal effect of teachers' quality on students' outcomes without concerns regarding the endogenous sorting of teachers and students.

## 4 Teacher Value Added Measures

### 4.1 How Do We Compute TVA?

The dataset we use to estimate TVA is stacked as a panel. The observation unit is student-school-class-subject and year. TVA is computed uniquely for a teacher per year and grade. We estimate TVA following the procedure described by Chetty, Friedman, and Rockoff (2014a). The construction of TVA is implemented in three steps. First, student test scores are modeled as a function of student characteristics, and we compute test score residuals while adjusting for observables. These residuals contain the teacher's contributions to the student's test scores and some estimation noise. Second, the best linear predictor of the mean test score residuals for teachers in year t is estimated based on teachers' mean scores in all previous and later years. Third, each teacher's average residuals at each point in time are predicted using the average residuals of that same teacher at every other point in time. These predictions are the final TVA measures. We implement this using Michael Stepner's vam Stata program with minor adjustments for our setting. For each teacher, these TVA estimates are the best linear predictions of teachers' value added in each observed year, based on the scores of students taught by that teacher in other years (prior and future). Each teacher's VA is not assumed to be fixed over time. We apply a drift adjustment, since (a) test scores from more recent classes better predict current teacher quality, and (b) this adjustment improves out-of-sample TVA forecasts (Chetty, Friedman, and Rockoff, 2014a).

We included in the sample teachers who had taught a given grade in a school for at least 2 years. Therefore, TVA cannot be estimated for teachers who appear in our sample for only 1 year.[21] To alleviate any concern that our estimates are biased due to teachers potentially teaching the same students across grades, we eliminate students who have the same teacher in two consecutive grades. For the 2003-2005 sample, there were 110 such teachers. In the TVA regressions, we use baseline control variables similar to those of Chetty, Friedman, and Rockoff (2014a) and Kane, Rockoff, and Staiger (2008). These include students' demographics

[21] 172 teachers teach for only 1 year in the sample and 144 teachers teach in all years in the sample.

(gender and age), student high school track indicators, quadratic polynomials of a student's previous-year test scores in the same subject, class-level means of prior test score in the same subject, class size, school-level-grade enrollment, gender of the teacher, teacher's experience, means of prior-year average GPA, class and school-grade means of prior-year average GPA, average prior test scores in the same subject, school's neighborhood income, grade, and year FE. When a prior test score (1.3% of our sample) is missing, we set it to 0 and include an indicator for missing data. Teacher experience is based on the teacher's workload during the study period—namely, how many yearly classes and courses a teacher taught during the study period.

We compute TVA for 936 teacher-grade-year configurations. We use data from 341 unique classrooms and around 48,000 stacked observations. We scale TVA in standard deviations of the distribution of student test scores. TVA in our sample varies widely from x-2.8 to 2.8. This range is similar to findings in the literature, which confirms that teachers vary substantially in effectiveness (Hanushek, 2011; Hanushek and Rivkin, 2010). The distribution of TVA is presented in Figure 2, and we notice a substantial variation in teacher VA. In Figure A2, we show the distribution by gender of teachers. There are no noticeable differences between male and female teachers. The differences in mean TVA by teacher gender presented in Table 1 show a similar pattern. Figure A3 presents the distributions of TVA for teachers in grades 11 and 12 grades separately. No noticeable differences are observed between the two.

For the 2003-2011 sample, we similarly compute TVA, but we split this period into three equal periods: 2003-2005, 2006-2008, and 2009-2011. We compute the average TVA of a teacher in each of these three subperiods using only $12^{th}$-grade data. We do account for drift in teacher quality in these subgroups. We use the same set of controls (except grade FE) to calculate TVA in the period 2003-2011 as in 2003-2005.[22]

### 4.2 How Stable Is Teacher VA Across Classes in Different Grades?

We use the multiple TVA estimates for each teacher to examine their stability within teachers over grades and years. Table 4 presents correlation coefficients between these multiple TVA estimates for a teacher. We use TVA estimates obtained based on the 2003-2005 sample in Panel A. The correlation between TVA in $11^{th}$ and $12^{th}$ grades is 0.491 when we condition on school and year fixed effects (column 1) and 0.489 when we add controls for teacher gender (column 2). Then we examine the correlation of within-teacher VA estimates over time. When all controls are added, the correlation between TVA measured in 2004 and 2003 is 0.785, in 2004 and 2005 is 0.813, and in 2003 and 2005 is 0.915. Teacher VA is computed using both grades from 2003-2005 in columns (3)-(8). All of these estimates are statistically different from 0 at the 1% significance level. Panel A establishes that we observe high within-teacher correlations between TVA measured in both

[22] A histogram of the TVA distribution in $12^{th}$ grade using the 2003-2011 data is shown in Figure A4.

grades in 2003-2005.

We show additional evidence in Panel B while including the 2006-2011 data. For this period, we can compute TVA only for $12^{th}$-grade teachers because the national exams in $11^{th}$ grade were discontinued in 2006. We also recalculate the TVA in 2003-2005 using only $12^{th}$ grade data for consistency. The correlation between a teacher's VA in 2003-2005 and 2006-2008 is 0.821. The estimated correlation between TVA measures in 2003-2005 and 2009-2011 is lower 0.448 and between 2006-2009 and 2009-2011 is 0.571. These TVA estimates suggest high persistency in TVA across time and grades.

## 5 Alternative Measures of TVA

Our previous definition of TVA measures high school teachers' effectiveness in improving their students' test scores at the end of the grade. In line with recent studies in the literature, we test the hypothesis that teachers may have effects on long-term outcomes that go beyond students' short-term tests (Gilraine and Nolan, 2021; Rose, Schellenberg, and Shem-Tov, 2022; Petek and Pope, 2023). To do so, we use a similar methodology as in Section 4.1, but replace the (residualized) end-of-grade test scores with the (residualized) longer-term student-level outcomes in the measurement of TVA. In particular, we use (a) the percentile rank of the enrolled degree quality and (b) a student's university admissions score as an outcome in the measurement of TVA. A higher degree quality rank is associated with a university degree in which students with a higher university admissions score enroll. High long-run TVA teachers are teachers who significantly contribute to students' academic growth over an extended period of time, and not only end-of-grade test scores. Intuitively, this long-term TVA measure captures the long-run effect of being randomly assigned to a teacher who effectively improves long-term outcomes.

Are teachers who help students perform better on exams the same teachers who help students be successful in their postsecondary admissions? We examine the correlations between test-score TVA and long-run TVA. We also examine the correlations between different long-run value added measures. Then, we construct longer-run TVA measures and examine the impact of being assigned to long-run value added high school teachers on university admissions scores, quality of enrolled degree at the university level, more preferred study programs, postsecondary admissions, enrollment in an academic university compared with a technical institution, and the likelihood of winning a state government scholarship.

We then assess value added measures of teacher quality that capture disruptive student behavior in the class to examine the behavioral impact of teachers on students. To do so, we follow Jackson (2018) and Petek and Pope (2023) and replace (residualized) test score outcomes with (residualized) unexcused absences of students in the following year. In our case, this behavioral value added measure only captures a teacher's ability to reduce a student's propensity to be suspended from the class. Noncognitive value added in the other studies captures log days absent (plus one), GPA, effort GPA, indicator for getting suspended, and an indicator for not progressing to the next grade on time (i.e., held back).

## 6 Effect of Teachers' Quality on Short- and Long-term Outcomes

We estimate the following specification to assess the impact of a teacher's quality on students' high school and postsecondary schooling outcomes:

$$Outcomes_{i,s,c,g,j,p,t} = \alpha_u + \gamma X_{i,s,c,t} + \beta X_{j,s,c,t} + \delta TeacherVA_{j,s,c,t,g,p} + s_s + \psi_c + \zeta_g + \chi_p + u_t + k_i + \epsilon_{i,s,c,g,j,p,t}, \tag{1}$$

where $Outcomes_{i,s,c,g,j,p,t}$ denotes the outcome of student $i$ in school $s$, class $c$, and grade $g$, assigned to teacher $j$, in subject $p$ and year $t$; $s_s$ is a school fixed effect; $\psi_c$ is a class fixed effect; $\zeta_g$ is a grade fixed effect; $\chi_p$ is a subject fixed effect; and $u_t$ is a year fixed effect. We also add a $k_i$ student fixed effect in the most demanding specification. $TeacherVA_{j,s,c,t,g,p}$ is a measure of a teacher $j$'s quality measured by their value added in school $s$, class $c$, year $t$, grade $g$, and subject $p$. The TVA is scaled in units of student test-score standard deviations. The controls in equation (1) include student characteristics $X_{i,s,c,t}$ and teacher characteristics $X_{j,s,c,t}$. Student characteristics include gender and age. Teacher characteristics include gender and teacher experience. We cluster standard errors at the school-by-cohort level to account for the fact that students face common school-by-cohort-level shocks (Chetty, Friedman, and Rockoff, 2014a,b). The coefficient of interest is $\delta$ and captures the effect of a teacher's quality on a student's subsequent performance in the same subject. We also estimate the effect of TVA on a series of longer-term outcome variables. Our longer-term outcomes are student-specific, but do not vary by grade, subject, or year. Thus, we modified specification (1) so that all variables were aggregated at student level. The variable of interest is then the average TVA a student was exposed to in the senior year of high school. Relevant outcomes include the university admissions score, the quality rank of an enrolled postsecondary degree, the rank of the institution attended on a student's degree preference list, a 0/1 indicator for enrollment in a university versus a technical school, a 0/1 indicator for enrollment in postsecondary schooling, and a 0/1 indicator for whether students won a state government scholarship to pursue their postsecondary studies.[23] We apply two correction methods in the above estimation. First, we implement an empirical Bayes (EB) shrinkage estimation approach to address potential sampling error, because the TVA estimates may be based on small samples for some teachers.[24] In the second strategy,

[23] For student-level analyses, we cluster standard errors at class level, although our patterns of results are similar under different clustering options.

[24] This method constructs an unbiased measure of TVA that accounts for noise in the measurement (Lavy and Sand, 2018; Lavy and Megalokonomou, 2024; Terrier, 2020). In particular, the noisy measure of a teacher VA is multiplied by an estimate of

we use a two-step bootstrapping technique to correct for the fact that the main variable of interest (TVA measure) is a generated regressor.[25]

# 7 Main Results

## 7.1 Short-term Test-score TVA

We first present a simple graphical presentation of the relationship between TVA and student test scores. The upper panel in Figure 3 presents a binned scatter plot of test scores in year $t$ and TVA. The relationship is positive, and the bins are concentrated around the diagonal line. This positively sloped relationship indicates that changes in the teaching staff's quality strongly predict test-score changes. In contrast, when we use student gender (=1 if female) or student age on the y-axis of the binned scatter plot, we see that TVA changes do not predict changes in female share or students' age. The upper panel is a visual representation of the main result, and the lower panel is a visual representation in which students' characteristics are uncorrelated with teacher quality. Given the random match between teachers and students in Greece, the latter result is expected.

In Table 5, we present estimates of TVA's effect on students' high school outcomes.[26] Column 1 presents TVA estimates for $11^{th}$-grade national exam test scores, and column 2 for $12^{th}$-grade scores in 2003-2005. In both columns, the specification includes controls for teacher's gender and experience, student's subject-specific prior test scores, and year, track, subject, class, and student fixed effects. We can use this within-student estimation because each student is tested in several subjects taught by a different teacher. The two estimates are positive and statistically significant at the 1% level. Standard errors are clustered by school and year. Estimates in columns 1-2 show that a 1-standard-deviation (SD) improvement in TVA raises test scores by approximately 0.20 SD in $11^{th}$ grade and 0.21 SD in $12^{th}$ grade. In column 3, we pool the $11^{th}$- and $12^{th}$-grade test scores and run it with a sample of 42,731 observations in one regression. The estimate we obtained, 0.205, is in the midrange of the grade-specific estimates. Column 4 presents the TVA estimates

its reliability. This reliability term is the ratio of signal variance to signal variance plus noise variance. In this way, less reliable measures of TVA (those with large variations in estimated residuals) are shrunk back toward the mean of the distribution of the TVA measure.

[25]This procedure is performed in two steps. Two-step estimations obtain inconsistent standard errors in the second-stage regressions, as they don't account for the presence of a generated regressor (Pagan, 1984). We follow a two-step bootstrapping method to compute standard errors (Ashraf and Galor, 2013; Lavy and Megalokonomou, 2024; Terrier, 2020). The bootstrap estimates of the standard errors are constructed as follows: In the first stage, a random sample of students is drawn with replacements from each teacher's classes. Then a new measure of TVA is calculated using equation (1) based on the new sample of students. In the second stage, we estimate the effect of this newly created TVA on student test scores and other outcomes, and the coefficients are saved. The bootstrapped random samples are used in both parts of the procedure. The two-step bootstrap sampling is repeated 1,000 times. The standard deviations in the sample of 1,000 observations of coefficient estimates from the second step are the bootstrapped standard errors of the estimated effects of teacher quality.

[26]We exclude students who had the same teacher in two consecutive grades in the same subject from the analysis. This is the case for 10% of students.

for $12^{th}$-grade national exam test scores using data from 2003 to 2011. The estimated effect is very similar to the one we obtained in the 2003-2005 period for the $12^{th}$ grade. A 1 SD improvement in TVA raises test scores by 0.192 SD in $12^{th}$ grade.

These estimated effects of TVA on test scores are larger than those of Chetty, Friedman, and Rockoff (2014a), based on grades 4-8 in a primary school in a large urban school district in the U.S. Their estimate is 0.14 SD for teacher VA's effect on math test scores and 0.10 SD for English test scores. Bau and Das (2020) use data from Pakistan and find that a 1 SD increase in TVA raises test scores in math by 0.21 SD and English by 0.17 SD. Our context is high schools, and the test scores are from high-stakes national exams because they are used for admission to postsecondary schooling. Another difference is that we pool many subjects in our sample, not only math and English.

In Table A12, we use two correction techniques to estimate the impact of short-term test-score TVA on subsequent subject-specific test scores. We present estimates using EB shrinkage estimations and the two-way bootstrapping methods in Table A12. These estimated effects are statistically significant and remain similar to those presented in Table 5. We view these findings as supporting evidence of the credibility of the method we use to measure TVA quality and the causal interpretation of our findings.

We then show the effects of teacher VA based on test scores on six longer-term outcomes. These outcomes are measured at student level and not student-by-subject level. Therefore, we use the average TVA of a student's teachers across all subjects in a given academic year as the variable of interest. We used the longer period from 2003 to 2011. Therefore, our sample includes one observation for each student-school cell, and the regressions do not include subject or student-fixed effects. We present these estimates in Panel A in Table 6. Column 1 shows the effect of test-score TVA on a student's university admissions score. A 1 SD increase in test-score TVA improves students' university admissions score by 1.762 relative to a mean of 12.312. Descriptive information for all variables used here can be found in online Appendix Table A2. Column 2 in Table 6 presents the effect of TVA on the quality of the postsecondary institutions in which students enroll. The estimated effects show that a 1 SD increase in TVA increases the quality of the institution a student attends by 14.171 rank points when quality is measured based on the marginal test score cutoff for admission.[27] This finding is consistent with Chetty, Friedman, and Rockoff (2014a), who find that higher TVA causes students to attend better colleges. Their measure of college quality is the average earnings of older graduates. High school teacher VA's positive effect is also evident in the likelihood of enrolling in a more desired major: A 1 SD higher teacher VA pushes students almost 3 places up on their list of preferred study programs relative to a mean of 9.790. Column 4 shows that a 1 SD increase in TVA increases the likelihood of

[27] The effect is very similar if we measure postsecondary degree quality by the average admissions score of all enrolled students in each college degree.

attending postsecondary schooling by 0.113 percentage points. Column 5 shows that a 1 SD higher test-score TVA is also associated with a 0.142-percentage-point higher likelihood of attending an academic university than a higher education technical school. Column 6 indicates that there is also a 0.03 percentage point higher likelihood of winning a state government scholarship upon enrollment in a postsecondary degree when they were assigned to higher-TVA high school teachers. All of these estimates are measured precisely.

### 7.2 Long Term Outcomes TVA

So far, we have presented evidence that teacher effectiveness measured by short-term test-score TVA affects students' short- and longer-term outcomes. We now derive the long-run TVA measures. We first present a histogram of the short-term test-score TVA and the two different long-run VAs we measured using students' long-run outcomes in the measurement of TVA. In Figure 4, we show the related histograms: The upper panel shows the distributions of the short-term test-score TVA and the TVA based on degree quality. The bottom panel presents the distributions of the short-term test-score TVA and the long-run TVA based on student university admissions scores. The traditional test-score TVA exhibits a larger dispersion of values, while both long-run outcomes TVA measures show a distinctive concentration of TVA values in the (-1,1) range. Figure 4 indicates that there is a substantial overlap between the two long-run TVA measures.

We then examine the impact of long-run TVA on students' longer-term outcomes. Panels B and C in Table 6 show the estimated effects of long-run TVA on students' longer-term outcomes. In Panel B, the treatment variable is the average TVA based on the degree quality a student was admitted to. In Panel C, the treatment variable is TVA measured based on the student's university admissions score. Most estimated effects in Panels B and C are positive and statistically indistinguishable from zero. This indicates that students assigned to higher long-run TVA teachers succeed substantially more in long-term outcomes. Panel B indicates that a 1-SD-increase in long-run TVA pushes students almost 4 places up on their list of preferred study programs. A 1 SD higher TVA renders students more likely to attend postsecondary schooling by 0.406 percentage points and to win a merit scholarship for outstanding performance from the State Government Scholarship by 0.127 percentage points. The estimates in Panel C are smaller than those in Panel B on average, but still larger compared with the test-score TVA estimates shown in Panel A. These estimates reveal that effective long-term TVA teachers contribute significantly to students' academic growth over an extended period.

We then explore the correlation between the short-term test-score TVA and each of the long-run TVA measures. Table 7 shows these correlations for the entire sample and different subsamples. The sample now includes one observation for each teacher-year configuration. We notice that teachers who are effective in improving students' test scores are also likely to increase their longer-term outcomes. In the entire sample, the correlation is 0.77 and 0.88 in the teacher FE and year FE specification in column (2) in Panels A and B, respectively. The correlation only slightly varies among subgroups: male teachers, female teachers, teachers with experience below the median, and teachers with teaching experience above the median. The correlations vary from 0.43 to 0.94. This evidence indicates that teachers who are effective in improving test scores are, on average, also effective in improving long-term outcomes. Table A16 in the online Appendix presents the correlations between long-run TVA measured based on degree quality and long-run TVA measured based on university admissions score. The correlation between these two long-run VAs is around 0.5.

### 7.3 Behavioral TVA Measure

We then measure a behavioral TVA by using students' unexcused absences at the end of grade 12. Effective teachers help students reduce their unexcused absences or be less prone to be suspended. Thus, the expected direction of the effects would generate negative estimated effects on the various outcomes. Table A17 shows the impact of our behavioral TVA on students' longer-term outcomes. Indeed, teachers who are effective in reducing suspension improve students' university admissions scores. Effective teachers also increase students' likelihood of enrolling in an academic university compared with a technical postsecondary school and winning a merit scholarship for outstanding performance. However, the correlations between short-term test-score TVA and behavioral TVA is zero.[28] This indicates that teachers who are highly effective in raising students' test scores are not necessarily the same teachers who are effective in improving student behavior. Our finding of a smaller correlation between short-term test-score TVA and behavioral TVA is in line with other studies (Gilraine and Nolan, 2021). However, other studies find a higher correlation between long-run TVA and behavioral TVA. Perhaps our finding is different here because our behavioral TVA measure is based on fewer aspects of student behavior compared with other studies (Gilraine and Nolan, 2021; Jackson, 2014).

### 7.4 Simultaneous Controls for Short-term Test-score TVA, Long-run TVA, and Behavioral TVA Measures

Table A18 shows the estimated effects when we simultaneously include the test-score TVA, long-run TVA, and behavioral TVA in the same regression. The objective of this exercise is to ascertain which TVA measure predominates in predicting student long-term outcomes and examine whether these TVA measures exert independent effects on long-term outcomes. In Panel A, the long-run TVA is based on the quality of the enrolled degree; in Panel B, it is based on national exam performance. Test-score TVA estimates are always

[28] We also find that zero correlation between long-run TVA and behavioral TVA in Appendix Table A19.

statistically significant, but lower when we also control for long-run TVA and the behavioral TVA in the same regression compared with the main results in Table 5. This implies that some of the effect of test-score TVA captures the effect of the quality of university degree TVA.

Also, the estimated effects of the quality of university degree TVA in Table A18 are much lower than those in Table 6, and only one remains statistically significant. Regarding the estimated effects' magnitudes, both test-score TVA and long-run TVA appear to be effective in predicting students' long-term success, although the latter demonstrates lower precision. Behavioral TVA also explains students' longer-term outcomes in the expected direction, and 4/6 of the estimated coefficients are statistically significant. A similar pattern arises in Panel B. Short-term test-score TVA, longer-term TVA (based on national exam performance), and behavioral TVA all explain students' longer-term outcomes. The test-score TVA estimates are similar to those in Panel A and remain consistently statistically significant. Long-run TVA estimates are less precise overall compared with test-score TVA estimates. In the two cases in which long-run TVA estimates are statistically significant, they are even larger than the respective test-score TVA estimates. The estimated effect of long-run TVA on postsecondary degree quality is 13.814 (se=6.504), while the respective estimated effect of test-score TVA is equal to 12.265 (se=1.833). The effects of behavioral TVA are at the majority, statistically indistinguishable from zero. These effects indicate that teachers who effectively reduce student suspensions also contribute to better long-term student outcomes, even after accounting for other short-term test-score TVA and longer-term TVA. The magnitude of the behavioral TVA coefficients are smaller than that of short-term test-score TVA and longer-term TVA.

Our findings are broadly consistent with those of Gilraine and Nolan (2021), who found that long-run TVA provides the best prediction of teachers' effects on students' long-term success. However, in our study, these effects exhibit lower precision. This variance may stem from the different definitions of long-run TVA. Gilraine and Nolan (2021) define long-run TVA using subsequent subject-specific test scores. In our case, TVA is defined based on students' quality of enrolled degree or overall national exam score on university admissions exams. Thus, our analysis in Appendix Table A18 is at student level and involves only one observation per student, leading to a small sample size (=2,136). Their long-run TVA likely captures teachers' enduring impacts on subsequent subject-specific test scores, perhaps indicative of a deeper comprehension of subject material. In contrast, our long-run TVA is measured by students' choices regarding postsecondary admission, including the quality of their enrolled degree and overall performance on university admissions exams.

### 7.5 Robustness

Next, we present results that support our identification strategy. In Table A13, we test how sensitive the main test-score TVA effects on subsequent test scores shown in Table 5 are to the specification we used. We focus on the estimates presented in columns 3 and 4 of Table 5 derived from the stacked data of $11^{th}$ and $12^{th}$-grade test scores (2003-2005) and the $12^{th}$ grade (2003-2011) only.

In Table A13, in the first row, we re-estimate the regression that generated column 3 in Table 5. Still, we start from the simplest specification and gradually add controls (teacher characteristics, class FE, and student FE). The full specification, presented in column 4 of Table A13, is identical to the specification used in column 3 of Table 5. In the next row, we do the same with the specification that generated column 4 in Table 5. In this case, only $12^{th}$-grade data are used; thus, we do not include grade FE in the baseline controls. Estimates are generally mostly the same as we expand the specification. The estimate in column 1, based on the most minimal set of controls, is 0.196. When we add the teacher's gender and experience as controls, the estimate becomes 0.195. The standard error is unchanged, and both estimates are significant at the 1% level. In column 3, we replace the school fixed effect with a class fixed effect. The estimate increases from 0.195 (se=0.037) to 0.206 (se=0.033). After adding student fixed effects, the estimate is 0.192 (se=0.040). This is the full specification we also use in column 4 in Table 5. Note that 0.192 (in column 4) is not statistically different from 0.196 in column 1. The conclusion is that gradually adding controls to the regression does not change the estimates much.

In Table A14, we present a robustness exercise based on our Table 6, while focusing on the effects of multiple TVA measures on students' longer-term outcomes. Panels A, B, and C in Table A14 use the average TVA based on test scores, average TVA based on degree quality, and average TVA based on university admissions scores we used in Panels A, B, and C in Table 6, respectively. We again start from a basic specification, and as we move across columns we add controls until we reach the ones used in columns 1-6 in Table 6. The estimated effects remain very similar across columns and specifications. For example, TVA's estimated effect on the attended institution's rank changed from 2.223 (se=0.449) in column 1 to 2.779 (se=0.477) in column 4 in Panel A. The estimated effects of TVA based on degree quality on winning a state government scholarship change from 0.125 (se=0.045) in the simplest specification to 0.127 (se=0.045) when all controls are included. We believe that these robust results provide further evidence that teachers are randomly assigned to classes. As a result, TVA is not correlated with student classroom characteristics, which supports our findings' causal interpretation.

Next, we examine how robust the results presented in Table 5 are to adding other controls and changes in the different fixed effects specifications. These results are shown in Table A15. Column 1 in the top panel replicates the estimates in column 3, Panel A, of Table 5. In columns 2 and 3, we add to the basic specification of column 1 different combinations of class, year, grade, and student fixed effects to account for unobserved shocks that can confound our estimates. Such unobserved shocks can lead to a mechanical but spurious relationship between TVA and students' performance. In column 2 we include class-by-subject-by-grade fixed effects, and in column 3 we include grade-by-student-by-year fixed effects to account for shocks that may be specific by grade by student and by year. In column 4, we include the same controls as in column 3 in Table 5 but replace class fixed effects with school fixed effects. Column 5 presents estimates from a specification that consists of an indicator variable for core subjects (versus track subjects). The impact estimate of TVA remains the same. The point estimate increases to 0.208 (se=0.020) from 0.205 (se=0.021). In column 6, we include school-by-grade-by-year fixed effects. Here we exploit the fact that we usually observe more than one teacher teaching in the same school within the same year in different grades, and thus we allow shocks to vary across years, grades, and schools. Again, the estimated effects and standard errors only change slightly. Remarkably, the point estimate of TVA's effect on test scores is very stable across all columns of Table A15. None of the variations in the specification make any difference. All columns' point estimates are between 0.205 and 0.212 and are highly statistically significant. These robust findings allow us to conclude that our main estimates are not sensitive to controls or unobserved shocks at school, grade, year, student level, or any combination.

The bottom panel in Table A15 refers to estimates derived using $12^{th}$-grade (2003-2011) national exam scores. For ease of comparison, we present again in column 1 the estimate from column 4 in Table A19. In column 2, we exploit the fact that we usually observe more than one teacher teaching the same student within the same class, and thus we allow shocks to vary across students and classes. In column 3, we include student-by-year fixed effects. In column 4, we include the same controls as in column 4 in Table 5, but we replace class fixed effects with school fixed effects. In column 5 we include an indicator variable for core subjects (versus track subjects), and in column 6 we include school-by-year fixed effects. The point estimate across specifications varies from 0.200 to 0.191 and is highly statistically significant. We conclude that TVA's estimated effects and the standard errors are very similar when we use different fixed effects.

### 7.6 Heterogeneity in the Effect of Teacher Value Added

In Table A20, we examine heterogeneity by student ability. We use test-score TVA in all regressions in the remaining sections. We allow for variation in the treatment effect by adding, in the main regression, an interaction term between TVA and students' previous-year test scores. Again, we note that these test scores from previous years are subject-specific. We also include a main effect for the previous year's test score. We present results from three specifications that vary by the controls we include in the regressions. TVA and previous-year test scores have a positive effect on current test scores. The estimated coefficient on the interaction term between TVA and previous-year test scores in all three specifications is positive and its precision increases as we add more controls. The within-student estimate of the interaction term is 0.040 (se=0.014). These results imply that TVA's positive effect increases with student ability, as measured by prior achievement.

In Tables A21 and A22 in the online appendix, we present estimates using separate samples of male and female teachers and separate samples of male and female students. Table A21 demonstrates that the estimated TVA's effect is positive and statistically significant for female and male teachers, and the two effect sizes are very similar. When we examine the $11^{th}$- and $12^{th}$-grade test scores, the estimates by teacher gender are virtually identical—0.188 for male teachers and 0.194 for female teachers. We have seen in Figure A2 that there are no gender differences in TVA, and here we find no differences in the effect of TVA by teacher gender. In Table A22, we present the estimated effects of TVA by the student's gender. The impact on boys and girls is very similar: The estimated effect for male students is 0.208 (se=0.025) and for female students 0.203 (se=0.027).[29]

### 7.7 Allowing for NonLinearity in the Effect of Treatment

To allow for a nonlinear effect of TVA on student outcomes, we compute quintiles of TVA and replace the single treatment in the regression with a set of quintile indicators. Table A31 presents estimates of the effects of switching to TVA's second, third, fourth, or fifth quintile (relative to the first quintile) on different test

[29] In Table A23, we look at the effect of TVA on students' test scores by type of subjects (classics, science, and exact science) in the core or track. When only core subjects are included (columns 1-3), all estimated effects are positive and significant, but there are no statistically significant differences across the different types of subjects. The estimated effects are positive and significant when science or exact science track subjects are included (columns 5-6). The impact of TVA is more pronounced in the exact science and science tracks than in the classics track. Figures A5 and A6 show the distributions of TVA for core subjects (Figure A5) and by high school track (Figure A6). In Table A24, we examine heterogeneity by student ability and type of subjects. We again allow for variation in the treatment effect by ability by adding in the main regression an interaction term between TVA and student previous-year test scores in classics core (column 1), science core (column 2), and exact science core (column 3) subjects. Columns 4-6 show estimated effects in classics track (column 4), science track (column 5), and exact science track (column 6) subjects. The estimated coefficient on the interaction term between TVA and the previous-year test score is positive in science, using both core and track subjects. The within-student estimate of the interaction term is 0.053 (se=0.020) for science core subjects (column 2) and 0.059 (se=0.032) for science track subjects (column 5). These results imply that TVA's positive effect is increasing with students' ability in science subjects as measured by prior achievement. The interaction term for classics track subjects is negative and statistically significant, while for exact science core subjects the estimated effects are small and insignificant. In Table A25, we look at the effect of TVA on test scores for male and female students for different core subjects. There are no remarkable gender differences by type of subject. In Table A26, we focus on the effect of TVA on test scores for combinations of male and female students and teachers for different subjects. TVA's estimated effects are more pronounced when teachers and students are females than when both teachers and students are male in science and exact science subjects. In Table A27, we look at the effect of TVA on test scores for male and female students for different track subjects. Again, TVA's estimated effects seem more pronounced when teachers and students are female than when both teachers and students are male in science subjects. Figures A7 and A8 show the histograms of TVA for different subjects and for male and female teachers separately. Figure A7 focuses on core subjects, while Figure A8 shows the distributions of TVA for each high school track separately. There are no noticeable differences by teacher gender. Also, in Tables A28 and A29, we find no heterogeneity concerning teacher experience and teacher gender or class size and teacher gender, respectively.

scores. The first quintile is the bottom quintile of teacher quality, and higher quintiles are of higher quality. The effects appear to increase with quintiles for all subjects. Column (1) shows the pooled effect of TVA on test scores in all subjects, while columns (2)-(3) and (4)-(5) present the effects of TVA on test scores in classics and science core and track subjects, respectively. Columns (6)-(7) show the estimated effects of TVA on test scores in all classics and science subjects separately. The effect of TVA also increases in column (6). Focusing on column (7), students assigned to the second, third, fourth, and fifth quintiles in TVA (compared with the bottom quintile) improve their test scores by 0.07, 0.10, 0.15, and 0.23 SD, respectively. Overall, it seems that the estimated effects of TVA increase when moving from lower to higher quintiles of TVA for all subjects.

### 7.8 Does TVA in High School Affect the Choice of University Major?

In Table 8, we present the estimated effects of TVA on students' choices regarding their field of study at the university. The dependent variable is the decision to study in one of the following university departments or groups: economics, business, biology, history, mathematics, physics, engineering, computer science, health-related fields (medicine, dentistry, veterinary, and pharmacy); the remaining Humanities departments; remaining Science departments; and remaining Exact Science departments.[30] We estimate three specifications based on the controls used and results in Table 8.

Teacher value added is calculated as the average TVA in the closest high school subject to the student's university field of study. The average track TVA is used whenever there is no exact subject correspondence. We stack possible postsecondary choices as the dependent variable for each student against TVA in the closest high school subject. Since in the data several rows correspond to the same student, the number of observations is much larger than in previous tables. The dependent variable is a 0/1 indicator, assuming a value of 1 for the observed department of study and a value of 0 for other possible choices. We use the following high school subjects for each field of study: For economics, we use the TVA in economics in $12^{th}$ grade (otherwise, the TVA in the track in $12^{th}$ grade). For business, we use the TVA in business administration in the track in $12^{th}$ grade (otherwise, the TVA in mathematics in the core in $12^{th}$ grade). For history, we use the TVA in history in the track in $12^{th}$ grade (otherwise, the TVA in history in the core in $12^{th}$ grade). For mathematics, we use the TVA in mathematics in the track in $12^{th}$ grade (otherwise, the TVA in mathematics in the core in $12^{th}$ grade). For physics, we use the average TVA in physics in the track in $12^{th}$ grade (otherwise, the TVA in physics in the core in $12^{th}$ grade). For engineering, we use the average TVA in physics and biology

[30]The remaining humanities departments include degrees in archaeology, sociology, music studies, theatrical studies, religious studies, and Balkan studies. The remaining exact science departments include degrees in material technology, technology and innovation, biosystems engineering, food technology, multimedia design, applied information science in economics, geographical information systems and topology, environmental studies, and commercial navy engineering. The remaining science departments include agricultural science, genetics, geology, statistics, and neuroscience degrees.

in the track in $12^{th}$ grade (otherwise, the average TVA in physics and biology in the core in $12^{th}$ grade). For computer science, we use the TVA in computer science in the track in $12^{th}$ grade. For health-related fields (medicine, dentistry, veterinary, and pharmacy), we use the average TVA in the science or exact science track in $12^{th}$ grade (otherwise, the TVA in mathematics in the core in $12^{th}$ grade). For the remaining humanities departments, we use the average TVA in the classics track in $12^{th}$ grade (otherwise, the average TVA in modern Greek and history in the core in $12^{th}$ grade). For the remaining exact science departments, we use the average TVA in the exact science track in $12^{th}$ grade (otherwise, the average TVA in mathematics and physics in the core in $12^{th}$ grade). For the remaining science departments, we use the TVA in biology in the science track in $12^{th}$ grade (otherwise, the average TVA in mathematics and physics in the core in $12^{th}$ grade).

In the first row, we show the effects using the full sample (females and males), focusing on females in Panel B and males in Panel C. The absolute size of TVA's estimated effect is positive and statistically different from zero across all specifications for the full sample. The estimated effect is 0.034 (se=0.003) when the baseline controls and fixed effects are included in column 1, and it remains the same when more controls are added. In column 2, we also include student and average teacher characteristics; in column 3, we also include the average class size. This estimate means that a 1 SD increase in TVA in the closest high school subjects increases the probability that students are more likely to choose a major in this general field of study by 3.4%. The estimated effects for female and male students are positive, of similar magnitude, and statistically significant, which suggests that both genders react similarly to changes in TVA in their university study choice.

Given concern regarding the low enrollment of women in STEM higher education studies, we further explore the effect of TVA on the choice of field of study by focusing our analysis on students in science and exact science tracks in high school. In column 4 of Table 8, we show results based on the sample that includes students in the science and exact science high school tracks. We use the same specification as in column 3. Estimates in column 4 show that almost all of the effect on the choice among science and exact science students is on females (0.015), with almost no effect on males (0.008). In particular, TVA's estimated effect on females' field of study in STEM tracks is 0.015 (se=0.007) and has practically no effect on males (estimated effect equals 0.008 with se=0.006). It seems that more productive teachers (higher TVA) not only improve female achievements in STEM subjects in high school but motivate more of them to choose a STEM field of study in higher education. Such an effect is not observed for men. This is a significant result, because it implies that improving teachers' test scores' value added will reduce the gender imbalance in university schooling's STEM areas.

In Table A30, we present the estimated effects of TVA for the closest high school subjects on the students' likelihood to enroll in a university department that is a natural follow-up of the high school track. The outcome variable is a binary indicator that takes the value 1 if students enroll in a university field (exact science, science, humanities, social science) equivalent to their high school track (exact science, science, humanities). We group fields of study based on four broad study tracks. Humanities include liberal arts, literature, psychology, journalism, philosophy, education, Greek language, history, foreign languages, home economics, and law. Social sciences include economics, business and management, accounting, political science, and European studies. Exact science includes mathematics, engineering, physics, and computer science. Science includes biology, chemistry, medicine, pharmacy, veterinary studies, and dentistry. Again, we model students' choices using linear probability regression. The average TVA in the related high school track is used for each possible choice of the dependent variable.[31] The dependent variable is a binary indicator that assumes the value 1 for the observed field of study and 0 for the other three choices.

As described earlier in this section, TVA is calculated as the average TVA in the closest high school subjects to a student's university department of study. The track average is used whenever there is no exact subject correspondence. In Table A30, TVA's estimated positive effect remains almost unchanged across the three specifications. The estimated effect is 0.045 (se=0.021) in column (3), which implies that a 1 SD increase in TVA in the closest subject in high school increases the probability that students will choose a related university department as their field of study by 4.5 percentage points. Estimates for female and male students are presented in the second and third rows. For female students, the estimated effect is 0.036 in column (3), which is the most augmented specification. The estimated effect for male students is positive (0.055 in column 3), and marginally significant at the 10% level across specifications. Tables 8 and A30 show evidence that female and male students do not react differently to teacher VA changes in their choice of study at university level.

We present a similar analysis in column 4 Table A30 while limiting the samples to students from science and exact science high school tracks. The estimated effects of teacher VA is similar for females and males in science and exact science study areas.

### 7.9 Mechanisms Behind The Effect of Teacher Value Added: Student Attendance

In this section, we examine whether TVA affects test scores through time of instruction in the classroom. We expect that students exposed to less instruction time will learn less and have lower test scores (Lavy, 2015;

[31] We use the average track TVA in the exact science high school track for social science university degrees, since most students who enroll in social science degrees follow the exact science track in high school.

Lavy and Megalokonomou, 2024; Lavy, 2020).[32] We measure variation in students' exposure to teachers' instruction using data on students' absenteeism throughout the year. We observe two types of absenteeism, excused and unexcused, during regular school days. Both are measured in terms of hours of instruction. Absenteeism is available for each student but does not vary by subject. Therefore, to examine the effect of hours absent on regular school days, we use a sample that includes one observation for each student in a given grade—i.e., a student-grade-year-school cell.

Excused absenteeism refers to occasions on which a student knows in advance that they cannot attend school on a specific date or time because of an obligation to be elsewhere—for example, a medical appointment or family emergency. Often, the school grants permission for such absenteeism in advance or it is authorized by parents, usually with a note signed by a doctor or a parent regarding a short-term illness. Therefore, such absenteeism can be correlated with unobservables that are correlated with school performance and exam test scores. Unexcused absenteeism refers to occasions when a student misses some lessons or full school days without getting permission from some school authority. This reflects a student's choice not to attend school on a particular day. Absence due to a student's suspension is also reported as unexcused absenteeism.

We have information on students' absenteeism by type for both grades, 11 and 12. In Table 9, we report estimates for three outcomes: total absence, excused absence, and unexcused absence. In column (1), we present estimates based on a specification that includes school-by-grade-by-year fixed effects, previous-year test scores, track fixed effects, and controls for lagged attendance. We ran two more specifications: one that includes student characteristics and one that adds teacher characteristics and class-by-cohort characteristics. TVA's estimated effects on total absences are negative but with a large standard error. This remains the same across the different specifications. TVA's estimated effects on excused absences are positive and much smaller—practically almost zero—but very imprecise. TVA's effect on unexcused absences is negative, large, and significantly different from zero. In the baseline specification, the estimate is -3.509 (se=1.172). When student characteristics are added, the estimate declines to -2.892 (se=1.172). Adding teachers' characteristics and school-by-cohort characteristics has practically no effect on the point estimate (-2.847, se=1.132). The estimated coefficients' size implies that a 1 SD improvement in TVA reduces unexcused absences by 3 hours. This is a considerable effect compared with the mean number of unexcused absences per year per student, which is 27 hours. Note that the three specifications' estimates are not statistically different, which implies that adding the controls in columns 8 and 9 does not change the point estimate. These results suggest that TVA's effect on students' cognitive performance on national exams is partly mediated by increasing or

[32] There is evidence that high-quality teachers help students improve complex cognitive skills and social-emotional competencies (Kraft, 2019); reduce grade retention and suspensions and improve high-school completion (Jackson, 2018; Koedel, 2008) and student self-reported attitudes and behaviors in class (Blazar and Kraft, 2017).

decreasing students' exposure to teaching time in the classroom. The mechanism that affects this exposure is absenteeism on regular school days or from classes, whereby students miss material covered in class. This evidence also suggests that unexcused absences are not correlated with student, teacher, class, or cohort characteristics.

Table A32, column 2 shows that unexcused absenteeism is uncorrelated with student characteristics such as age and gender. However, unexcused absenteeism is a proxy for student motivation and student disruption in the classroom. We further support the claim that students with lower motivation are less likely to attend classes by showing that students with lower previous-year test scores have more unexcused absences.

In Table A33, we examine the effect of absenteeism on student performance by type of subject. We use seven groups of subjects: all subjects, classics in the core, classics in the track, science in the core, science in the track, exact science in the core, and exact science in the track. We include teacher characteristics, previous-year test scores, controls for TVA, and subject, year, grade, student, and class fixed effects in all specifications. If anything, we find a negative relationship between student absenteeism and performance. When we look at the effect of total absences on performance in Panel A, the estimates are small and insignificant. However, an additional hour of absence reduces performance by -0.003 of a SD when we focus on exact science track subjects. The pattern is similar when we focus on the impact of excused absences on performance in Panel B. In Panel C, we find that an increase in unexcused absences will likely reduce student performance in classics track subjects. This relationship is imperfect, since the absenteeism data are not at the subject level, while student performance is at the subject level. Nevertheless, this table indicates a negative correlation between missing class and performance.

# 8 Teacher Value Added and Teacher Turnover in Schools

## 8.1 Teacher Value Added and the Entry and Exit of Teachers

Since teacher quality determines school quality, a critical question is which schools retain high-productivity teachers and attract new ones with high potential value added. Is this factor the dividing line between high- and low-performing schools? This section examines teachers' entry and exit and how it relates to their TVA. The data we use in this paper allow us to identify these mobility dynamics in our sample during the study period, with some limitations. Regarding teachers who leave their school, we do not know their destination—whether they moved to another school, took a non-teaching job, or changed to non-employment. For teachers who entered a school in our sample during the study period, we do not know whether they are novice teachers, moved from another school, or came from another occupation or non-employment. Subject to these constraints, our aim in this section is to examine the correlational relationship between TVA and the entry and exit of teachers and determine whether they vary by teacher gender and the school's quality.

Table A34 presents descriptive statistics for the entry and exit of teachers in our sample. Since we use the whole sample period (2003-2011) in this part of the analysis, we use data only for the $12^{th}$ grade. The sample includes 1,267 teacher-year cells; 27% were hired during the study period. We refer to them as "new" teachers. An equal proportion exited their schools during the same period. The average tenure of teachers in the sample is 4.8 years; the maximum is 9, and the minimum is 2. TVA measures are computed for 927 teacher-year cells because we cannot measure TVA for exiting teachers.

We computed TVA for each teacher and year they are observed in the sample. We then examine TVA's effect on each teacher's exit probability the following year. We use these data for the years 2004-2011. TVA is again scaled in units of standard deviations of the student-level test score distribution. The data we use in this estimation include all classes taught by the same teacher in all 16 other classes and years in the sample.[33] The outcome variable is a teacher's retention indicator, which is equal to 1 if the teacher stayed in school for the following year and 0 otherwise.

There is no difference in the mean TVA of new and exiting teachers in the sample (Panel B of Appendix Table A34). The lowest TVA range for each group is also very similar for the two groups of teachers. However, the mean TVA of teachers who leave their school during the study period is lower than all teachers' mean TVA (-0.116 compared with -0.167). This means that our sample's teachers who exit schools are among the lowest-TVA teachers. Panel C of Table A34 presents similar statistics for high- and low-achieving schools. We use the school-average test score in all subjects in 2003 to rank schools and divide the sample in high- and low-performing schools—using the median as a cutoff. Remarkably, teachers' mean entry and exit rates in the two subsamples are very similar: The entry rate is 0.27 in both groups, and the exit rate is 0.28 and 0.26 in high- and low-achieving schools, respectively. However, the average TVA in high-achieving schools is higher by 0.12 SD than in low-achieving schools. In the rest of this section, we examine the nature of the relationship between teachers' entry/exit and their TVA. In particular, we aim to determine whether teachers who leave their schools are among the best or worse regarding TVA.

In Table 10 columns (1)-(3), we present estimates from a regression of the probability that a teacher remains in their school in year t+1 on TVA conditional on teaching at the school in year t. We run this regression for all teachers in the school in t-1 and for teachers who joined the school in year t. We use three specifications: a regression that includes as controls school and year fixed effects, a second that includes teachers' gender, and a third in which we add as a control previous-year test scores of the students assigned to each teacher. The first row and columns (1)-(3) show that TVA negatively correlates with teachers'

[33]Table 1 shows that students take 16.5 subjects on average in $11^{th}$ and $12^{th}$ grades.

likelihood to remain in the same school—namely, better quality teachers are less likely to remain in the school. The estimate is robust across the three specifications. Its size, 0.045, means that a 1 SD increase in TVA reduces the likelihood of remaining in the school by 4.5% (or equivalently, a 1 SD increase in TVA increases the likelihood of exiting the school by 4.5%). In the lower panel, we present the estimated TVA effects for teachers who joined the school a year before. This effect is doubled in magnitude (mean=0.10) and statistically different from zero for new teachers who joined the school the year before. The estimates from these regressions, which are robust to changes in specification, show that a 1 SD increase in TVA will lower a teacher's tenure in that school beyond year $t$ by 10%.

## 8.2 Which Schools Do Good Teachers Leave?

In Table 11, we replicate the above analysis by stratifying the sample above and below the school's median quality.[34] Lower-quality schools are responsible for these mobility dynamics, whereby good teachers are less likely to remain in those schools (equivalently, a higher probability of leaving). The estimate obtained for low-achieving schools (below median quality) is -0.064 in the most augmented specification and significant at the 10% level. The estimate based on schools above the median quality is much smaller (-0.034) and significantly different from -0.064. The same pattern is observed in the lower panel for the retention likelihood of new teachers. The estimated effect of TVA for schools above the median quality is 0.062 (se=0.053) in column 3. The estimate for schools below the median quality is -0.117 (se=0.053) in column 6 and significant at the 5% level. It remains very robust to different specifications. The results indicate that low-achieving schools observe many of their high-productivity teachers leave every year, while this is not the case among high-achieving schools. This is even more the case for new high-quality teachers who exit low-quality schools soon after they join them. These findings are consistent with the following potential explanation: High-quality teachers stay in high-performing schools and prefer not to leave because of better working conditions such as a pleasant environment, good peers, more benefits, etc.

Table 10 reports evidence on the relationship between TVA and retention decisions for female (columns 4-5) and male teachers (columns 6-7) separately. Male teachers drive the effects. The estimate for male teachers in column (7) is -0.092 (se=0.036), which indicates that a 1 SD increase in TVA reduces the probability of retention by 9 percentage points for male teachers. The related estimated effect for female teachers is practically zero (estimated coefficient in column (5)=-0.024 with se=0.029). The same pattern holds for new teachers. The estimated effect for new male teachers is more than double in magnitude compared with all teachers. In particular, it is equal to -0.212 and statistically significant at the 1% level. This indicates that a 1 SD increase in TVA reduces the probability of retention by 21% for new male teachers. The estimated

[34]In Figure 6, we plot the distributions of TVA in low-quality and high-quality schools.

effects are practically zero for new female teachers (estimated coefficient=-0.018 with se=0.053). This is the case, although we have already shown (Figure A2) that female and male teachers have, on average, the same TVA.

The evidence that male teachers tend to leave for other schools more than female teachers is consistent with evidence on gender differences in job mobility in other settings in the labor market. For instance, Theodossiou and Zangelidis (2009) find substantial gender differences in job transitions and show that men are more mobile across jobs than women using data from six European countries. This pattern is much more pronounced for workers with several years of experience. This may be consistent with our finding that the mobility effects are twice as large for new teachers, who are more likely to be fresh college graduates. However, we do not have data to test that directly. Using data from US full-time workers, Loprest (1992) shows that job mobility plays an important role in the wage growth of young men, who are more likely to change jobs than young females. Among academic economists, Hilmer and Hilmer (2010) find that the number of job moves is positively associated with wages for men in academics, while the corresponding estimated effect is small and not statistically significant for women. There is also evidence that men engage more in job searches while employed than women, possibly as a way to obtain a promotion and higher pay (Keith and McWilliams, 1995). Different mobility dynamics for females may be attributed to marital and family circumstances (Han and Moen, 1999).

# 9 Conclusion

This paper investigates how teacher quality affects students' high school performance and their enrollment in university schooling. We start by using the teacher's value added based on test scores to measure teacher quality. Then, we also use multiple measures of teacher quality, including long-run TVA measures and behavioral TVA. Long-run TVA captures teacher effectiveness in improving university schooling outcomes for students, and behavioral TVA measures teacher effectiveness in reducing suspensions. Using a unique education context in Greece, whereby high school teachers and students are assigned to classes that result in a random match between teachers and students, we can avoid concern that endogenous sorting of teachers and students could bias TVA measures.

We measure each teacher's standard test-score value added multiple times using teachers' complete teaching assignments over almost a decade. We find that high school students assigned to high-quality teachers have higher test scores on end-of-high-school national exams. These exams are used to determine admission to higher education in Greece, and therefore TVA affects postsecondary schooling through its effect on high school national exams. These effects translate to an impact of TVA on the likelihood of pursuing postsecondary education, particularly in universities—in terms of the quality of universities to which students are admitted—and the likelihood of students being admitted to their preferred study program. The effect size of these impacts is meaningful economically. For example, a 1 SD improvement in TVA raises normalized average test scores by approximately 0.20 SD in $11^{th}$ and $12^{th}$ grades. It also increases the likelihood of continuing university education by 0.10 percentage points. Remarkably, higher-TVA teachers in high school also affect the choice of field of study at university by increasing the probability of enrolling in an area related to the high school study track. These effects do not vary by the gender of the teacher. The richness of the data allows us to control for student-specific unobserved heterogeneity, which might vary by grade and year, and account for school-specific unobserved shocks, which may confound the estimates. Our results are robust to an extensive battery of robustness exercises.

We then measure the long-run value added of teachers using students longer-term outcomes, including university admissions-related outcomes. We find a substantial correlation between short-term test-score VA and long-run TVA measures. We also find that being randomly assigned to teachers with high long-run value added positively impacts student outcomes. We also measure value added based on teachers' ability to reduce students' suspensions. Although we find no correlation between short-term test-score or long-run TVA and behavioral TVA, we find that assignment to teachers who effectively reduce students' suspensions positively impacts student outcomes. We also examine how well short-term test-score TVA, long-run TVA, and behavioral TVA predict students' longer-term success when we include all of those TVA measures in the same regression. We find that all types of TVA contribute to predicting student long-term outcomes, with long-run TVA and short-term test-score TVA being the best predictors, although long-run TVA is less precise.

Higher TVA implies higher teaching quality, which results in better student learning. However, we identify school attendance as a second channel through which higher TVA affects students' learning and achievements. We find that higher TVA reduces school absenteeism. This increases the time exposure of students to teaching, and this augmented time input improves learning. As another potential mechanism, we examine whether schools use TVA to sort teachers in or out. We find a much higher exit rate from schools among higher-TVA teachers, and this pattern is only evident among low-performing schools. In higher-achieving schools, the exit rate of high-TVA teachers is almost zero. We also find that high-performing schools hire teachers who soon show higher TVA than new teachers in low-performing schools. These findings suggest that observed TVA and unobserved potential TVA are used in determining the entry and exit of teachers in good schools and less so in low-performing schools.

## References


Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics 25*(1), 95–135.

Ashraf, Q. and O. Galor (2013). The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *American Economic Review 103*(1), 1–46.

Bacher-Hicks, A., M. J. Chin, T. J. Kane, and D. O. Staiger (2017). An Evaluation of Bias in Three Measures of Teacher Quality: Value-Added, Classroom Observations, and Student Surveys. *National Bureau of Economic Research, Working Paper No. 23478*.

Baker, E. L., P. E. Barton, L. Darling-Hammond, E. Haertel, H. F. Ladd, R. L. Linn, D. Ravitch, R. Rothstein, R. J. Shavelson, and L. A. Shepard (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. EPI Briefing Paper# 278. *Economic Policy Institute*.

Bau, N. and J. Das (2020). Teacher Value Added in a Low-Income Country. *American Economic Journal: Economic Policy 12*(1), 62–96.

Bizopoulou, A., R. Megalokonomou, and S. Simion (2022). Do Second Chances Pay Off? *CESifo Working Paper No. 9620*.

Blazar, D. and M. A. Kraft (2017). Teacher and Teaching Effects on Students? Attitudes and Behaviors,. *Educational Evaluation and Policy Analysis 39*(1), 146–170. PMID: 28931959.

Blazar, D., E. Litke, and J. Barmore (2016). What Does It Mean to be Ranked a "High" or "Low" Value-Added Teacher? Observing Differences in Instructional Quality Across Districts,. *American Educational Research Journal 53*(2), 324–359.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). 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.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources 41*(4), 778–820.

Corcoran, S. P., J. L. Jennings, and A. A. Beveridge (2011). Teacher Effectiveness on High- and Low-Stakes Tests. *Society for Research on Educational Effectiveness*.

Darling-Hammond, L. (2015). Can Value Added Add Value to Teacher Evaluation? *Educational Researcher 44*(2), 132–137.

Dinerstein, M., R. Megalokonomou, and C. Yannelis (2022). Human Capital Depreciation and Returns to Experience. *American Economic Review 112*(11), 3751–62.

Feld, J., N. Salamanca, and U. Zölitz (2020). Are Professors Worth It? The Value-Added and Costs of Tutorial Instructors. *Journal of Human Resources 55*(3), 836–863.

Gilraine, M. and P. Nolan (2021). Making Teaching Last: Long-Run Value-Added. Technical Report 29555, National Bureau of Economic Research.

Goulas, S., S. Griselda, and R. Megalokonomou (2022a). Comparative Advantage and Gender Gap in STEM. *Journal of Human Resources 0320-10781R2*.

Goulas, S., S. Griselda, and R. Megalokonomou (2022b). Compulsory Class Attendance versus Autonomy. *IZA Working Paper No. 14559*.

Goulas, S., B. Gunawardena, and R. Megalokonomou (2023). Same-Gender Peer Role Models in Education. *Mimeo*.

Goulas, S. and R. Megalokonomou (2018). Which Degrees do Students Prefer During Recessions? *Empirical Economics 56*(6), 2093–2125.

Goulas, S., R. Megalokonomou, and Y. Zhang (2022a). Female Neighbors, Test Scores, and Careers. *CESifo Working Paper, No. 10112*.

Goulas, S., R. Megalokonomou, and Y. Zhang (2022b). Females, Classroom Disruption and STEM Careers. *Working Paper*.

Green III, P. C., B. D. Baker, and J. Oluwole (2012). The Legal and Policy Implications of Value-Added Teacher Assessment Policies. *Brigham Young University Education and Law Journal*, 1.

Han, S.-K. and P. Moen (1999). Work and Family Over Time: A Life Course Approach. *The ANNALS of the American Academy of Political and Social Science 562*(1), 98–110.

Hanushek, E. A. (2011). The Economic Value of Higher Teacher Quality. *Economics of Education Review 30*(3), 466–479.

Hanushek, E. A. and S. G. Rivkin (2010). Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review 100*(2), 267–71.

Hilmer, C. and M. Hilmer (2010, May). Are there gender differences in the job mobility patterns of academic economists? *American Economic Review 100*(2), 353–57.

Horvath, H. (2015). Classroom Assignment Policies and Implications for Teacher Value-Added Estimation. *Unpublished Manuscript*.

Jackson, C. K. (2014). Teacher Quality at the High School Level: The Importance of Accounting for Tracks. *Journal of Labor Economics 32*(4), 645–684.

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.

Kane, T., D. Staiger, and W. Seattle (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains (Policy and Practice Brief). Bill and Melinda Gates Foundation.

Kane, T. J., D. F. McCaffrey, T. Miller, and D. O. Staiger (2013). Have we Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. *Research Paper. MET Project. Bill & Melinda Gates Foundation*.

Kane, T. J., J. E. Rockoff, and D. O. Staiger (2008). What Does Certification Tell Us About Teacher Effectiveness? Evidence From New York City. *Economics of Education Review 27*(6), 615–631.

Kane, T. J. and D. O. Staiger (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. *National Bureau of Economic Research, Working Paper No. 14607*.

Keith, K. and A. McWilliams (1995). The Wage Effects of Cumulative Job Mobility. *ILR Review 49*(1), 121–137.

Kinsler, J. (2012). Assessing Rothstein's Critique of Teacher Value-Added Models. *Quantitative Economics 3*(2), 333–362.

Koedel, C. (2008). Teacher Quality and Dropout Outcomes in a Large, Urban School District. *Journal of Urban Economics 64*(3), 560–572.

Koedel, C. and J. R. Betts (2011). Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique. *Education Finance and Policy 6*(1), 18–42.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015). Value-Added Modeling: A Review. *Economics of Education Review 47*, 180–195.

Kraft, M. A. (2019). Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies. *Journal of Human Resources 54*(1), 1–36.

Lavy, V. (2013). Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments Among Teachers. *The Economic Journal 123*(569), 540–573.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *The Economic Journal 125*(588), F397–F424.

Lavy, V. (2020). Expanding School Resources and Increasing Time on Task: Effects on Students? Academic and Noncognitive Outcomes. *Journal of the European Economic Association 18*(1), 232–265.

Lavy, V. and R. Megalokonomou (2024). The Short- and the Long-Run Impact of Gender-Biased Teachers. *American Economic Journal: Applied Economics 16*(2).

Lavy, V. and E. Sand (2018). On the Origins of Gender Gaps in Human Capital: Short- and Long-term Consequences of Teachers' Biases. *Journal of Public Economics 167*(C), 263–279.

Lockwood, J., D. F. McCaffrey, L. S. Hamilton, B. Stecher, V.-N. Le, and J. F. Martinez (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement 44*(1), 47–67.

Loprest, P. J. (1992). Gender Differences in Wage Growth and Job Mobility. *American Economic Review 82*(2), 526–532.

McCaffrey, D. F., J. Lockwood, D. Koretz, T. A. Louis, and L. Hamilton (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics 29*(1), 67–101.

Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler (2000). How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings? *Journal of Policy Analysis and Management 19*(4), 547–568.

Newton, X. A., L. Darling-Hammond, E. Haertel, and E. Thomas (2010). Value-added Modeling of Teacher Effectiveness: An Exploration of Stability Across Models and Contexts. *Education Policy Analysis Archives 18*, 23.

OECD (2018). Education for a Bright Future in Greece. *Reviews of National Policies for Education, OECD. Publishing, Paris*.

Pagan, A. (1984). Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review 25*(1), 221–47.

Paufler, N. A. and A. Amrein-Beardsley (2014). The Random Assignment of Students into Elementary Classrooms: Implications for Value-Added Analyses and Interpretations. *American Educational Research Journal 51*(2), 328–362.

Petek, N. and N. G. Pope (2023). The Multidimensional Impact of Teachers on Students. *Journal of Political Economy 131*(4), 1057–1107.

Rose, E. K., J. T. Schellenberg, and Y. Shem-Tov (2022, July). The effects of teacher quality on adult criminal justice contact. Working Paper 30274, National Bureau of Economic Research.

Rothstein, J. (2009). Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables. *Education Finance and Policy 4*(4), 537–571.

Rothstein, J. (2010, 02). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement*. *Quarterly Journal of Economics 125*(1), 175–214.

Rothstein, J. (2017). Measuring the Impacts of Teachers: Comment. *American Economic Review 107*(6), 1656–1684.

Sass, T. R., J. Hannaway, Z. Xu, D. N. Figlio, and L. Feng (2012). Value Added of Teachers in High-Poverty Schools and Lower Poverty Schools. *Journal of Urban Economics 72*(2-3), 104–122.

Schochet, P. Z. and H. S. Chiang (2013). What are Error Rates for Classifying Teacher and School Performance Using Value-Added Models? *Journal of Educational and Behavioral Statistics 38*(2), 142–171.

Stacy, B., C. Guarino, and J. Wooldridge (2018). Does the Precision and Stability of Value-Added Estimates of Teacher Performance Depend on The Types of Students they Serve? *Economics of Education Review 64*, 50–74.

Stylianidou, F., G. Bagakis, and D. Stamovlasis (2004). Attracting, Developing and Retaining Effective Teachers. *OECD Activity, Country Background Report for Greece*.

Terrier, C. (2020). Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement. *Economics of Education Review 77*, 101981.

Theodossiou, I. and A. Zangelidis (2009). Should I Stay or Should I Go? The Effect of Gender, Education and Unemployment on Labour Market Transitions. *Labour Economics 16*(5), 566–577.

Table 1: Summary Statistics for Sample Used to Estimate Teacher Value Added Models, 2003-2005

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| ***Student Characteristics*** | | | | |
| Gender (1=Female) | 0.561 | 0.496 | 0 | 1 |
| Previous Year Test Scores (std) | -0.028 | 0.994 | -6.535 | 3.140 |
| Age | 17.262 | 0.824 | 16 | 42 |
| No. of Subjects per Student | 16.462 | 4.048 | 3 | 21 |
| ***Tracks of Specialization*** | | | | |
| Classics | 0.348 | 0.476 | 0 | 1 |
| Science | 0.235 | 0.424 | 0 | 1 |
| Exact Science | 0.415 | 0.493 | 0 | 1 |
| ***Class Characteristics*** | | | | |
| Class size | 20.606 | 6.321 | 1 | 37 |
| ***Students Outcomes*** | | | | |
| ***High School*** | | | | |
| Test Score (std) | -0.113 | 1.015 | -3.559 | 2.857 |
| ***University Enrollment*** | | | | |
| Post-secondary Schooling (0/1) | 0.809 | 0.393 | 0 | 1 |
| Academic University Vs Technical School (0/1) | 0.476 | 0.499 | 0 | 1 |
| Post-secondary Degree Quality (Rank 1-100) | 46.511 | 28.652 | 0 | 99.926 |
| Rank of Attending Institution on Degree Preference | 10.784 | 13.744 | 1 | 140.000 |
| ***Teacher Characteristics*** | | | | |
| Teacher VA (2003-2005) | -0.08 | 0.640 | -2.806 | 2.753 |
| Female Teachers | -0.074 | 0.674 | -2.806 | 2.753 |
| Male Teachers | -0.084 | 0.603 | -1.966 | 1.543 |
| Teacher's Gender (1=Female) | 0.518 | 0.500 | 0 | 1 |
| Teacher's Experience (based on Previous Workload) | 10.007 | 6.459 | 1 | 42 |

Notes: All statistics reported are for the sample used in estimating the baseline value added model (2003-2005). No. of subjects per student counts the total number of subjects studied in grades 11 and 12. This sample includes only non-missing previous-year test scores and other requisite controls to estimate the TVA model. Student data are from the administrative records of the sample of 21 schools in Greece described in the text. Test scores are standardized z-scores. Age is measured in years and on the day they take the $12^{th}$ grade exam. "Postsecondary Schooling" is a binary indicator that takes the value 1 if a student enrolls in postsecondary schooling and 0 otherwise. "Academic University vs Technical School" is a binary indicator that takes the value 1 if the enrolled postsecondary institution is an academic university and 0 if it is a technical school. "Postsecondary Degree Quality 1" is degree quality based on the university admissions score cutoff. "Postsecondary Degree Quality 2" is a degree quality based on enrolled students' annual mean national exam performance. "Rank of Attending Institution on Degree Preference" is the rank of the enrolled option on a student's preference list. The smaller this number, the more desirable this degree choice is for a student. Teacher characteristics are computed based on the teacher sample. Teacher value added estimates are teacher-, grade-, and year-specific. Student characteristics and outcomes are calculated in a stacked sample that has one observation per student-school-class-subject-year cell.

Table 2:
ADDITIONAL SUMMARY STATISTICS FOR SAMPLE USED TO ESTIMATE TEACHER VALUE ADDED MODELS, 2003-2005

| | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Schools | 21 | | | | |
| Classes | 339 | | | | |
| Teachers, full sample | 279 | | | | |
| Teachers in $11^{th}$ Grade | 147 | | | | |
| Teachers in $12^{th}$ Grade | 132 | | | | |
| Teachers with computed Teacher Value Added (TVA) | 279 | | | | |
| Teachers in $11^{th}$ Grade | 147 | | | | |
| Teachers in $12^{th}$ Grade | 132 | | | | |
| Students | 3,173 | | | | |
| Students in $11^{th}$ Grade | 2,870 | | | | |
| Students in $12^{th}$ Grade | 2,602 | | | | |
| Students in both Grades | 2,299 | | | | |
| Subjects in $11^{th}$ Grade, full sample | | 8.602 | 1.495 | 1 | 10 |
| Subjects in $12^{th}$ Grade, full sample | | 7.907 | 1.758 | 1 | 11 |
| Subjects in $11^{th}$ Grade with TVA | | 8.602 | 1.495 | 1 | 10 |
| Subjects in $12^{th}$ Grade with TVA | | 7.907 | 1.758 | 1 | 11 |
| Teachers per Student in $11^{th}$ Grade, full sample | | 5.225 | 1.302 | 1 | 9 |
| Teachers per Student in $12^{th}$ Grade, full sample | | 5.068 | 1.401 | 1 | 9 |
| Teachers per Student in $11^{th}$ Grade with computed TVA | | 5.225 | 1.302 | 1 | 9 |
| Teachers per Student in $12^{th}$ Grade with computed TVA | | 5.068 | 1.401 | 1 | 9 |
| TVA estimates | 895 | | | | |
| Stacked Observations, full sample | 42,734 | | | | |
| Stacked Observations with computed TVA | 42,734 | | | | |
| 2002-2003 | 13,379 | | | | |
| 2003-2004 | 14,624 | | | | |
| 2004-2005 | 14,731 | | | | |
| Stacked Observations with the Same Teacher in Previous Grade, in Same Subject | 0 | | | | |
| Teacher Experience (No of classes taught in the past), full sample | 895 | 10.007 | 6.459 | 1 | 42 |
| Teacher Experience (No of classes taught in the past) with computed TVA | 895 | 10.007 | 6.459 | 1 | 42 |
| Female Teacher, full sample | 279 | 0.541 | 0.499 | 0 | 1 |
| Female Teachers with computed TVA | 279 | 0.541 | 0.499 | 0 | 1 |
| Male Teacher, full sample | 279 | 0.459 | 0.499 | 0 | 1 |
| Male Teachers with computed TVA | 279 | 0.459 | 0.499 | 0 | 1 |

The dataset is stacked so that there is one observation for each student-school-class-subject-year cell. Teacher value added is calculated with respect to the teacher, year, and grade cell. The TVA is calculated for teachers who teach the same grade for at least 2 years. If a teacher teaches only 1 year or one grade, we cannot estimate their TVA. Teacher experience measures the previous workload in the study period. It calculates how many times (class-year-subject cell) a teacher has taught from 2003 to 2005. A teacher sample is used to compute the share of female teachers.

Table 3:
Balancing Test of Student Pre-Assignment Characteristics on Teacher Characteristics, 2003-2005

| | Student Characteristics | | | | |
|---|---|---|---|---|---|
| | GPA in $10^{th}$ Grade (1) | Mathematics in $10^{th}$ Grade (2) | English in $10^{th}$ Grade (3) | Gender (=1 Female) (4) | Age (5) |
| ***Teacher Characteristics*** | | | | | |
| Gender (=1 Female) | 0.006 | 0.004 | 0.002 | -0.003 | -0.001 |
| | (0.007) | (0.006) | (0.006) | (0.002) | (0.002) |
| *N* | 40,548 | 40,548 | 36,323 | 42,732 | 42,732 |
| Value Added$_t$ | 0.005 | 0.006 | 0.001 | 0.000 | -0.004 |
| | (0.008) | (0.007) | (0.008) | (0.003) | (0.003) |
| *N* | 40,548 | 40,548 | 36,323 | 42,732 | 42,732 |
| Value Added$_{t-1}$ | 0.011 | 0.010 | 0.006 | 0.001 | -0.002 |
| | (0.007) | (0.007) | (0.006) | (0.004) | (0.005) |
| *N* | 32,712 | 32,712 | 29,310 | 34,572 | 34,572 |
| Experience | 0.0004 | 0.0003 | 0.0002 | 0.0001 | 0.0001 |
| | (0.0002) | (0.0002) | (0.0003) | (0.0002) | (0.0001) |
| *N* | 40,548 | 40,548 | 36,323 | 42,732 | 42,732 |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each estimate in this table is generated by a different regression. The table reports OLS coefficients from separate regressions of each student's pre-assignment characteristics on each of the teacher's characteristics. Scores in $10^{th}$ grade (GPA, mathematics, and English) are standardized and have a zero mean and standard deviation of 1. We use the average standardized performance in algebra and geometry in in $10^{th}$ grade for mathematics. Students' pre-assigned characteristics include gender (=1 if female), previous-year test scores, and age. The dependent variable is the GPA in grade 10 in column (1), test scores in mathematics in grade 10 in column (2), test scores in English in grade 10 in column (3), a binary indicator for the gender of the student (=1 if female) in column (4), and age in column (5). Independent variables are listed vertically and include the respective teachers' characteristics. In particular, we use teacher gender (=1 if female), assigned teacher's previous-year quality (measured by the assigned teacher's previous-year value added), and teaching experience based on the previous workload. All regressions include controls for grade fixed effects, year fixed effects, track fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Within-Teacher Correlations between Teacher Value Added Measured in Different Grades and Time Periods

| **Panel A:** | Teacher VA 2003-2005 (12th Grade) | | Teacher VA 2003 | | Teacher VA 2005 | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| TVA in 2003-2005 (11th Grade) | 0.491 | 0.489 | | | | | | |
| | (0.054)*** | (0.053)*** | | | | | | |
| TVA in 2004 | | | 0.781 | 0.785 | 0.811 | 0.813 | | |
| | | | (0.017)*** | (0.017)*** | (0.017)*** | (0.017)*** | | |
| TVA in 2003 | | | | | | | 0.919 | 0.915 |
| | | | | | | | (0.025)*** | (0.025)*** |
| Observations | 467 | 467 | 549 | 549 | 561 | 561 | 527 | 527 |
| **Panel B:** | Teacher VA 2006-2008 (12th Grade) | | Teacher VA 2009-2011 (12th Grade) | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| TVA in 2003-2005 (12th Grade) | 0.821 | 0.821 | 0.478 | 0.448 | | | | |
| | (0.035)*** | (0.034)*** | (0.096)*** | (0.094)*** | | | | |
| TVA in 2006-2009 (12th Grade) | | | | | 0.564 | 0.571 | | |
| | | | | | (0.047)*** | (0.046)*** | | |
| Observations | 628 | 628 | 416 | 416 | 527 | 527 | | |
| School FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teacher Gender | | ✓ | | ✓ | | ✓ | | ✓ |

Notes: The dataset is stacked and we have one observation for each teacher-grade-year configuration. If the grade is specified, TVA is measured as average TVA across classes in the specific grade and year it is reported. If the grade is not specified, TVA is measured as the average teacher VA across classes in the year it is specified. For instance, TVA in 2003 was the average teacher VA across grades 11 and 12 in year 2003. Standard errors clustered at school and cohort levels are reported in parentheses. Data for all years in the dataset are used. In Panel A, both grades 11 and 12 are used in columns (3)-(8). In Panel B, only grade 12 is used.

Table 5: The Effect of Test-Score TVA on High School Outcomes

| | Subject Specific National Score $11^{th}$ Grade 2003-2005 | Subject Specific National Score $12^{th}$ Grade 2003-2005 | Subject Specific National Score, Stacked $11^{th}+12^{th}$ Grade 2003-2005 | Subject Specific National Score $12^{th}$ Grade 2003-2011 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| TVA based on Test Scores | 0.199 | 0.209 | 0.205 | 0.192 |
| | (0.026)*** | (0.036)*** | (0.021)*** | (0.040)*** |
| $N$ | 23,566 | 23,566 | 42,731 | 38,244 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | |
| Baseline FE (Year, Track, Subject, Class, Student FE) | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimates from an OLS regression, with standard errors clustered by school and cohort levels in parentheses. The treatment variable is the TVA of teachers based on students' subject-specific test scores. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so that we have one observation for each student-grade-subject-school-class-year cell in columns 1-4. In columns 1 and 2, the dependent variable is the student's standardized test z-score on $11^{th}$-grade and $12^{th}$-grade national exams in 2003-2005, respectively. In column 3, the dependent variable is the stacked student's standardized test z-score on $11^{th}$- and $12^{th}$-grade national exams in 2003-2005. In column 4, the dependent variable is the student's standardized test z-score on the $12^{th}$-grade national exam in 2003-2011. Student characteristics include age and gender; teacher characteristics include gender and experience based on previous workload in the 2003-2005 sample. Previous-Year Test Scores is a student's test score in the same subject in the previous grade. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 6: The Effect of Alternative TVA Measures on Longer-Term Outcomes

| | Panel A | | | | | |
|---|---|---|---|---|---|---|
| | University Admissions Score 2003-2011 | Postsecondary Degree Quality 2003-2011 | Rank of Attending Institution on Degree Preference 2003-2011 | Enrollment in Postsecondary Schooling (0/1) 2003-2011 | Academic University Vs Technical School (0/1) 2003-2011 | Winning a State Government Scholarship (0/1) 2003-2011 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Average TVA based on Test Scores | 1.762 | 14.171 | 2.779 | 0.113 | 0.142 | 0.033 |
| | (0.184)*** | (1.525)*** | (0.567)*** | (0.017)*** | (0.034)*** | (0.005)*** |
| | Panel B | | | | | |
| | University Admissions Score 2003-2011 | Postsecondary Degree Quality 2003-2011 | Rank of Attending Institution on Degree Preference 2003-2011 | Enrollment in Postsecondary Schooling (0/1) 2003-2011 | Academic University Vs Technical School (0/1) 2003-2011 | Winning a State Government Scholarship (0/1) 2003-2011 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Average TVA based on Degree Quality | 5.496 | 37.514 | 3.990 | 0.406 | 0.393 | 0.127 |
| | (0.770)*** | (5.452)*** | (2.226)* | (0.074)*** | (0.110)*** | (0.045)*** |
| | Panel C | | | | | |
| | University Admissions Score 2003-2011 | Postsecondary Degree Quality 2003-2011 | Rank of Attending Institution on Degree Preference 2003-2011 | Enrollment in Postsecondary Schooling (0/1) 2003-2011 | Academic University Vs Technical School (0/1) 2003-2011 | Winning a State Government Scholarship 2003-2011 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Average TVA based on University Admission Score | 2.805 | 20.811 | 2.138 | 0.268 | 0.218 | 0.054 |
| | (0.748)*** | (3.524)*** | (1.264) | (0.086)*** | (0.071)*** | (0.024)** |
| Student Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimates from an OLS regression, with standard errors clustered by class level in parentheses. In Panel A the treatment variable is the average VA of teachers based on test scores. In Panel B the treatment variable is the average TVA based on a long-term outcome–i.e., rank of degree quality. In Panel C the treatment variable is the average TVA based on another long-term outcome, i.e., university admissions score. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so that we have one observation for each student. Previous-Year Test Scores is $10^{th}$-grade GPA. The treatment variable in Panel A is average TVA based on the test scores a student is exposed to, while in Panel B the treatment variable is average TVA based on the degree quality a student is exposed to. All variables are measured in the 2003-2011 period. In column 1, the dependent variable is a student's university admissions score. The university admissions score is an average test score across all subjects students take university admissions exams on. In column 2, the dependent variable is a degree's quality based on the annual degree admissions cutoffs. In column 3, the dependent variable is the enrolled degree's ranking on the students' preference list. In column 4, the dependent variable is a binary indicator that takes the value 1 if a student enrolls in some postsecondary institution and 0 otherwise. In column 5, the dependent variable is a binary indicator for whether a student is admitted to an academic university vs a technical school. In column 6, the dependent variable is a binary indicator for whether a student receives a merit scholarship for outstanding performance from the State Government Scholarship. We report the estimated coefficient of TVA on the rank of the attending institution by reversing the regression sign. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 7: Relationship between Test-Score TVA and Each of the Long-Term Outcomes TVA

| Outcome: TVA based on Test Scores | | |
|---|---|---|
| **Panel A** | | |
| | All Teachers | |
| | (1) | (2) |
| TVA based on Rank of Enrolled Degree | 0.661 | 0.770 |
| | (0.208)*** | (0.214)*** |
| | Male Teachers | |
| TVA based on Rank of Enrolled Degree | 0.701 | 0.850 |
| | (0.234)*** | (0.231)*** |
| | Female Teachers | |
| TVA based on Rank of Enrolled Degree | 0.603 | 0.552 |
| | (0.307)* | (0.316)* |
| | Experience below Median | |
| TVA based on Rank of Enrolled Degree | 0.509 | 0.630 |
| | (0.223)** | (0.230)*** |
| | Experience above Median | |
| TVA based on Rank of Enrolled Degree | 0.639 | 0.427 |
| | (0.504) | (0.401) |
| **Panel B** | | |
| | All Teachers | |
| | (1) | (2) |
| TVA based on University Admissions Score | 0.856 | 0.880 |
| | (0.154)*** | (0.166)*** |
| | Male Teachers | |
| TVA based on University Admissions Score | 0.770 | 0.788 |
| | (0.148)*** | (0.165)*** |
| | Female Teachers | |
| TVA based on University Admissions Score | 0.977 | 0.943 |
| | (0.247)*** | (0.256)*** |
| | Experience above Median | |
| TVA based on University Admissions Score | 0.866 | 0.717 |
| | (0.273)*** | (0.387)* |
| | Experience above Median | |
| TVA based on University Admissions Score | 0.808 | 0.790 |
| | (0.199)*** | (0.199)*** |
| Teacher FE | ✓ | ✓ |
| Year FE | | ✓ |

Notes: Each row reports coefficients from separate OLS regressions of the outcome variable (test-score TVA) on TVA calculated based on the quality of students' enrolled degree program (Panel A) or TVA calculated based on students' university admissions score (Panel B). The unit of observation is at the teacher and year level. The median teacher experience is 7 classes. TVA is estimated using the baseline control vector described in the text. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. Standard errors clustered by school and cohorts are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 8: The Effect of Teacher Value Added in the Closest Subject on Choice of University Field of Study for Full Sample and By Gender, 2003-2011

| | Indicator for Choice of Field of Study at the University Level | | | |
|---|---|---|---|---|
| | All Tracks | | | Science and Exact Science Track |
| | (1) | (2) | (3) | (4) |
| Panel A: Full Sample | 0.034 | 0.034 | 0.034 | 0.012 |
| | (0.003)*** | (0.003)*** | (0.003)*** | (0.004)*** |
| *N* | 91,461 | 91,461 | 91,461 | 69,723 |
| Panel B: Females | 0.037 | 0.038 | 0.037 | 0.015 |
| | (0.003)*** | (0.003)*** | (0.003)*** | (0.007)** |
| *N* | 50,242 | 50,242 | 50,242 | 33,100 |
| Panel C: Males | 0.030 | 0.030 | 0.030 | 0.008 |
| | (0.004)*** | (0.004)*** | (0.004)*** | (0.006) |
| *N* | 41,219 | 41,219 | 41,219 | 36,623 |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| School FE | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | |
| Student Characteristics | | ✓ | ✓ | ✓ |
| Average Teacher Characteristics | | ✓ | ✓ | ✓ |
| Average Class Size | | | ✓ | ✓ |

Notes: The dependent variable in columns 1-3 is the decision to study in one of the following university departments or groups: economics, business, biology, history, mathematics, physics, engineering, computer science, health-related (medicine, dentistry, veterinary, and pharmacy); remaining Humanities departments; and remaining science departments and remaining exact science departments. The dependent variable in column 4 is the decision to study in one of the following university departments or groups: mathematics, physics, engineering, computer science, remaining science departments, and remaining exact science departments. TVA is calculated as the average TVA in the closest high school subjects to the student's university field of study. Whenever there is not an exact subject correspondence, the average track TVA is used. We stack the possible postsecondary choices as the dependent variable for each student against the TVA in each of the university studies' postsecondary choices. The dependent variable is a 0/1 indicator, assuming the value 1 for the observed department of study and 0 for other possible choices. We use the following subjects for each field of study: for economics, we use the TVA in economics in $12^{th}$ grade. For business, we use the TVA in business administration in the track in $12^{th}$ grade. For history, we use the TVA in history in the track in $12^{th}$ grade. For mathematics, we use the average TVA in mathematics in the track in $12^{th}$ grade. For physics, we use the average TVA in physics in the track in $12^{th}$ grade. For engineering, we use the average TVA in physics and biology in the track in $12^{th}$ grade. For computer science, we use the TVA in computer science in the track in $12^{th}$ grade. For health-related (medicine, dentistry, veterinary, and pharmacy), we use the average TVA in the science or exact science track in $12^{th}$ grade. We use the average TVA in the classics track in $12^{th}$ grade for the remaining humanities departments. We use the average TVA in the exact science track in $12^{th}$ grade for the remaining exact science departments. We use the TVA in biology in the science track in $12^{th}$ grade for the remaining science departments. *Student Characteristics* include controls for student age and gender (1= female). *Average Teacher Characteristics* include the share of female teachers in the related subjects and the average experience of teachers in the related subjects. *Previous-Year Test Scores* include a student's $10^{th}$ grade GPA. *, **, and *** denote significance at the 10%, 5% and 1% level.

Table 9: Relationship between Teacher Value Added and Student School Attendance

| | Total Absences | | | Excused Absences | | | Unexcused Absences | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Teacher VA | -0.937 | -1.747 | -2.270 | 2.572 | 1.145 | 0.576 | -3.509*** | -2.892** | -2.847** |
| | (2.654) | (2.607) | (2.500) | (2.255) | (2.136) | (2.084) | (1.172) | (1.172) | (1.132) |
| Obs. | 4,711 | 4,711 | 4,711 | 4,711 | 4,711 | 4,711 | 4,711 | 4,711 | 4,711 |
| School FE x Grade FE x Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Lagged Attendance | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student Characteristics | | ✓ | ✓ | | ✓ | | ✓ | ✓ | |
| Teacher Characteristics & School-Cohort Characteristics | | | ✓ | | | ✓ | | | ✓ |
| Mean Y | 52.04 | 52.04 | 52.04 | 25.12 | 25.12 | 25.12 | 26.92 | 26.92 | 26.92 |
| St. Dev Y | 34.66 | 34.66 | 34.66 | 23.81 | 23.81 | 23.81 | 16.62 | 16.62 | 16.62 |

Notes: The dependent variable is a student's total (excused and unexcused), excused, and unexcused absences measured in hours. Student excused absences are absences (measured in hours) resulting from a student's illness, and parents should obtain a medical certificate from a doctor. Student unexcused absences (measured in hours) are absences that result from a student's misbehavior or disruptiveness during class. The dataset includes information on students' grade-specific total, excused, and unexcused absences across subjects. Thus, *Teacher VA* here has been calculated as the average across-subject teacher quality a student was exposed to in a particular school-grade and year. *Student Characteristics* include a binary indicator for the gender of the student (=1 if female) and age. We also control for the student's previous-year test scores. *Teacher Characteristics* include the average across-subject share of female teachers and the average teacher experience a student is assigned to. *School-Cohort Characteristics* include controls for the average previous-year's test scores, class size, and share of female classmates. Standard errors are clustered at school and cohort levels and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 10: The Effect of Teacher Value Added at Time t on Teacher Retention Status in School in Year t+1 Full Sample and by Teacher Gender

| | Full Sample of Teachers | | | | | | |
|---|---|---|---|---|---|---|---|
| | All Teachers | | | Female Teachers | | Male Teachers | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Teacher VA | -0.045 | -0.044 | -0.046 | -0.024 | -0.024 | -0.086 | -0.092 |
| | (0.021)** | (0.021)** | (0.021)** | (0.029) | (0.029) | (0.036)** | (0.036)** |
| *N* | 960 | 960 | 960 | 465 | 465 | 462 | 462 |
| | Sample of New Teachers | | | | | | |
| | All New Teachers | | | Female New Teachers | | Male New Teachers | |
| Teacher VA | -0.099 | -0.050 | -0.095 | -0.013 | -0.018 | -0.192 | -0.212 |
| | (0.037)*** | (0.048) | (0.038)** | (0.052) | (0.053) | (0.050)*** | (0.051)*** |
| *N* | 342 | 342 | 342 | 176 | 176 | 167 | 167 |
| School FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teacher Gender | | ✓ | ✓ | | | | |
| Previous-Year Test Scores | | | ✓ | | ✓ | | ✓ |

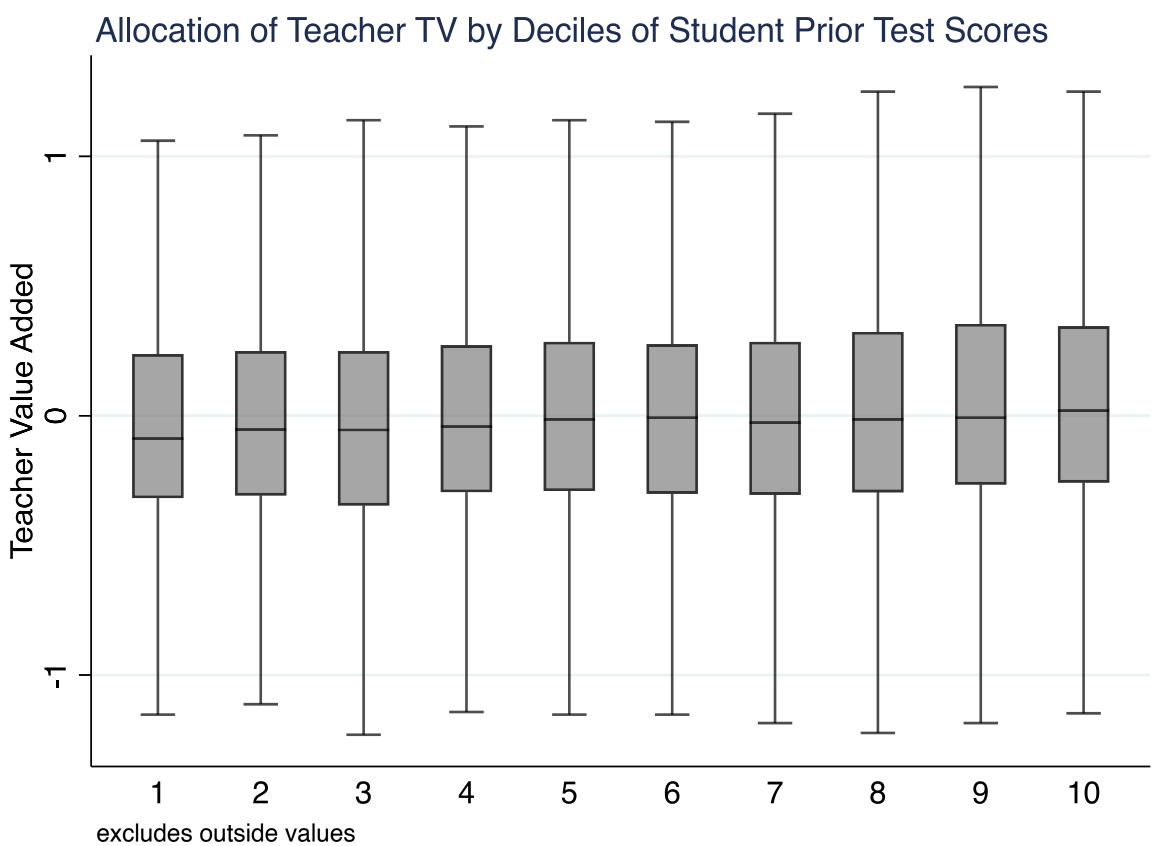
Notes: This table shows the estimated TVA effects on teacher retention status for all (columns 1-3), female (columns 4-5), and male (columns 6-7) teachers. Teacher retention status is a binary indicator that takes the value 1 if the teacher teaches in the next school year and 0 otherwise. Panel A includes all teachers, and Panel B only has new teachers. Teachers' TVA is measured in 2004, 2005, 2006, 2007, 2008, 2009, and 2010, and their retention status in 2005, 2006, 2007, 2008, 2009, 2010, and 2011. We pool TVA and retention status together and include a time dummy. A teacher teaches for up to 9 times in the same school in the data. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. Teacher Gender is a binary indicator that takes the value 1 if the assigned teacher is female and 0 otherwise. Previous-Year Test Scores measure each teachers' average previous-year performance of assigned students. Standard errors clustered at teacher level are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 11: The Effect of Teacher Value Added on Teacher Retention in School by Quality of School (Above and Below the Median)

| | All Teachers | | | | | |
|---|---|---|---|---|---|---|
| | Above Median Quality | | | Below Median Quality | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | -0.035 | -0.035 | -0.034 | -0.060 | -0.059 | -0.064 |
| | (0.026) | (0.026) | (0.027) | (0.034)* | (0.035)* | (0.035)* |
| *N* | 425 | 425 | 425 | 502 | 502 | 502 |
| | New Teachers | | | | | |
| Teacher VA | -0.071 | -0.065 | -0.062 | -0.115 | -0.116 | -0.117 |
| | (0.054) | (0.053) | (0.053) | (0.052)** | (0.053)** | (0.053)** |
| *N* | 157 | 157 | 157 | 186 | 186 | 186 |
| School FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teacher Gender | | ✓ | ✓ | | ✓ | ✓ |
| Previous-Year Test Scores | | | ✓ | | | ✓ |

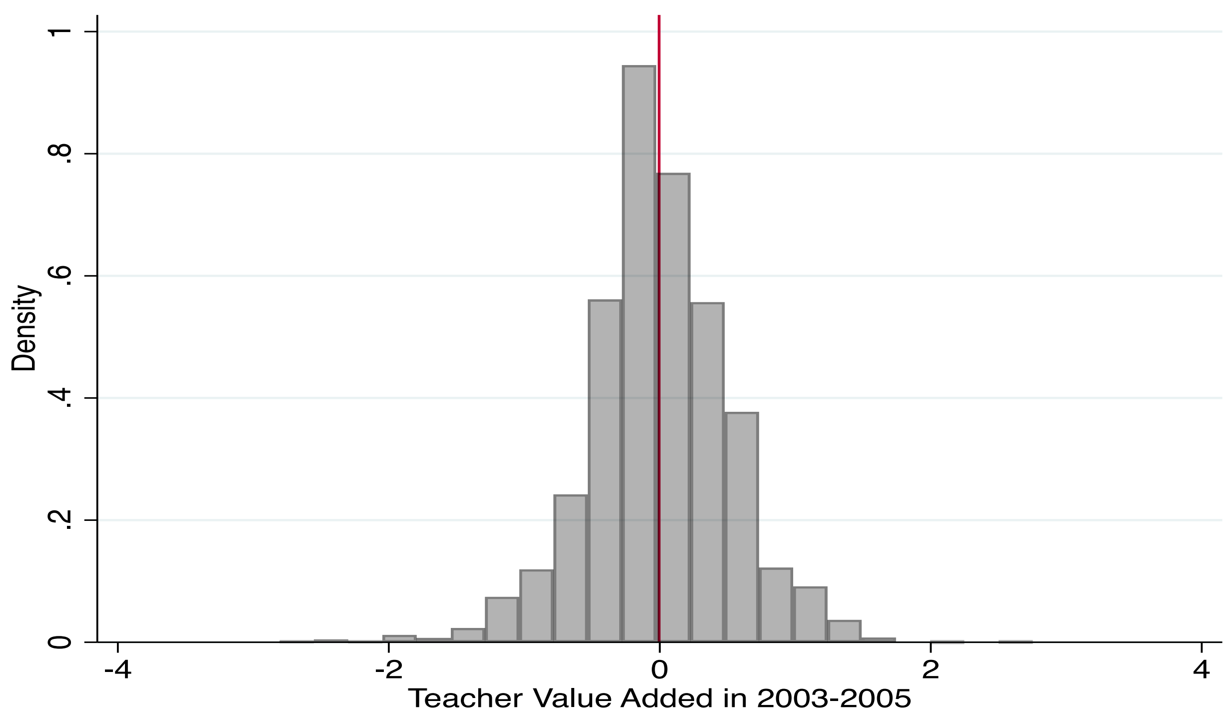
Notes: This table shows estimated TVA effects on teacher retention status for teachers in high- and low-quality schools. Teacher retention status is a binary indicator that takes the value 1 if the teacher teaches in the next school year and 0 otherwise. Panel A includes all teachers, and Panel B only has new teachers. Teachers' TVA is measured in 2004, 2005, 2006, 2007, 2008, 2009, and 2010, and their retention status in 2005, 2006, 2007, 2008, 2009, 2010, and 2011. We pool TVA and retention outcomes together and include a time dummy. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. Teacher Gender is a binary indicator taking the value 1 if the assigned teacher is a female and 0 otherwise. *Previous-Year Test Scores* measure each teacher's average previous-year test scores of assigned students. School Quality is calculated based on a ranking of the mean performance on national university entrance exams of students who attended each school in 2003, the first year in the sample. Only one school out of 21 did not operate in 2003 but opened in 2004. For this school, we used its school rank in 2004 to determine its school quality. We then determine whether a school's quality is above or below the median. Standard errors clustered at teacher level are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Figure 1:
Variation of Teacher Value Added by Decile of Student Previous-Year Test Scores (Additional Evidence of Randomization)



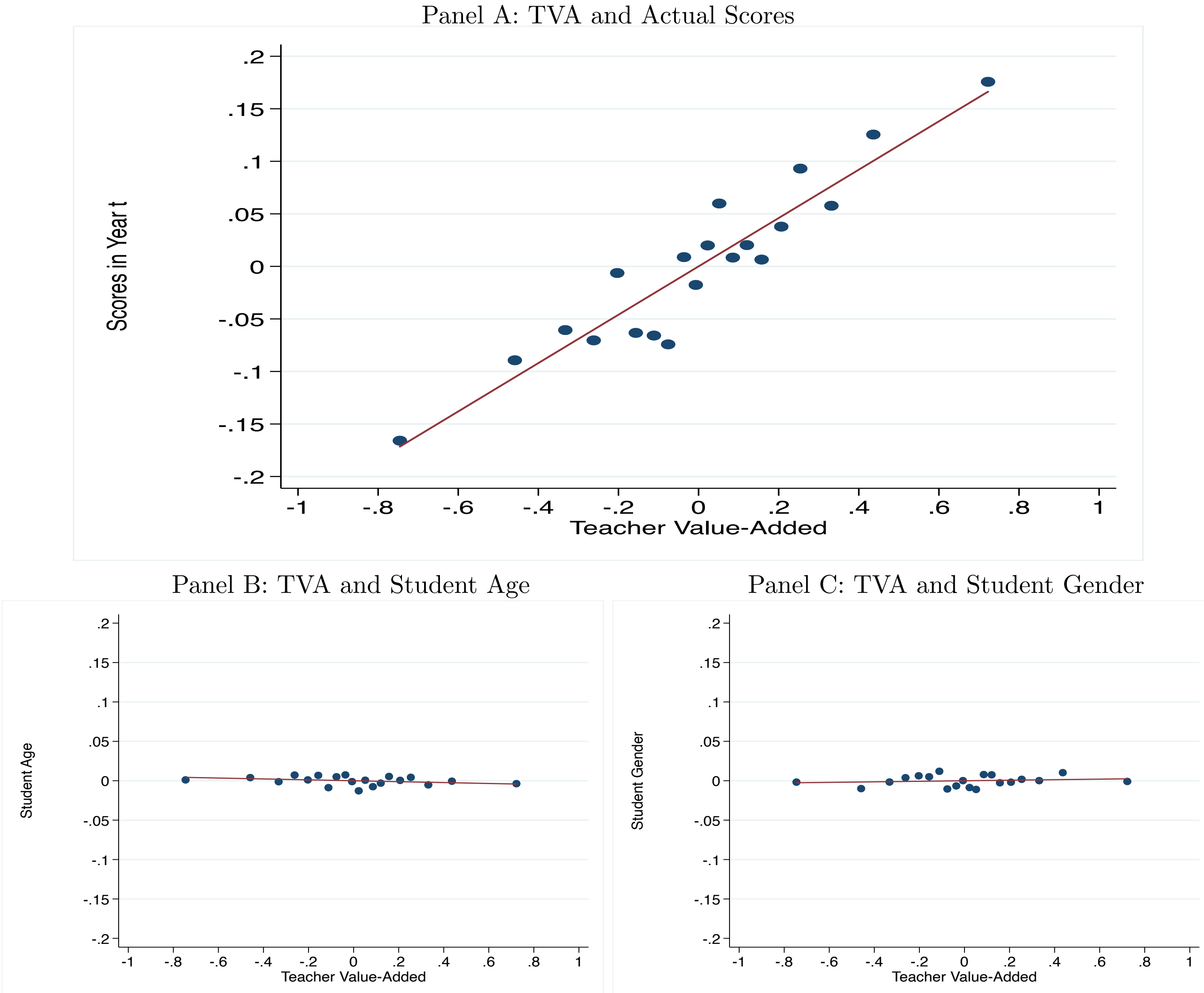

Notes: We split students' previous-year test scores into 10 deciles and compute the distribution of the TVA of students' assigned teachers for each decile of prior performance. We use the 2003-2005 data for the $11^{th}$- and $12^{th}$-grade sample to do this. Students in all different prior performance deciles are assigned to teachers of similar quality, on average, and the variation in TVA is very similar across deciles. This figure provides additional evidence that students with prior high or low performance are not more likely to be assigned to high- or low-quality teachers, which would violate teacher random assignment to students.

Figure 2:
HISTOGRAM FOR TEACHER VALUE ADDED MEASURE USED IN THE ANALYSIS




Notes: This histogram presents the distribution of the TVA measure based on test scores. To derive these value added measures, we pool the $11^{th}$- and $12^{th}$-grade data for the years 2003-2005, and we use $10^{th}$- and $11^{th}$- grade performance as previous-year test scores, respectively. This sample includes only students with non-missing previous-year test scores and other requisite controls to estimate the TVA model. Teacher VA is estimated using the baseline control vector, which includes previous-year own-subject test scores, student-level characteristics including age, gender, a binary indicator for being born in the first quarter of the birth year, class size, school-grade enrollment and school-grade and year-dummies. The structure of the dataset is one observation per teacher-year-grade combination.

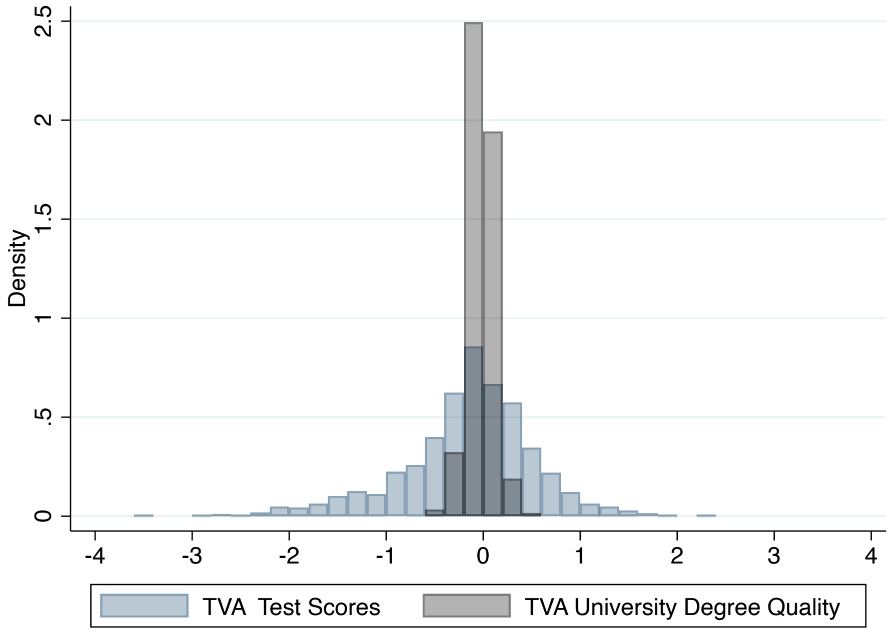
Figure 3:
CORRELATION OF TEACHER VALUE ADDED AND TEST SCORES



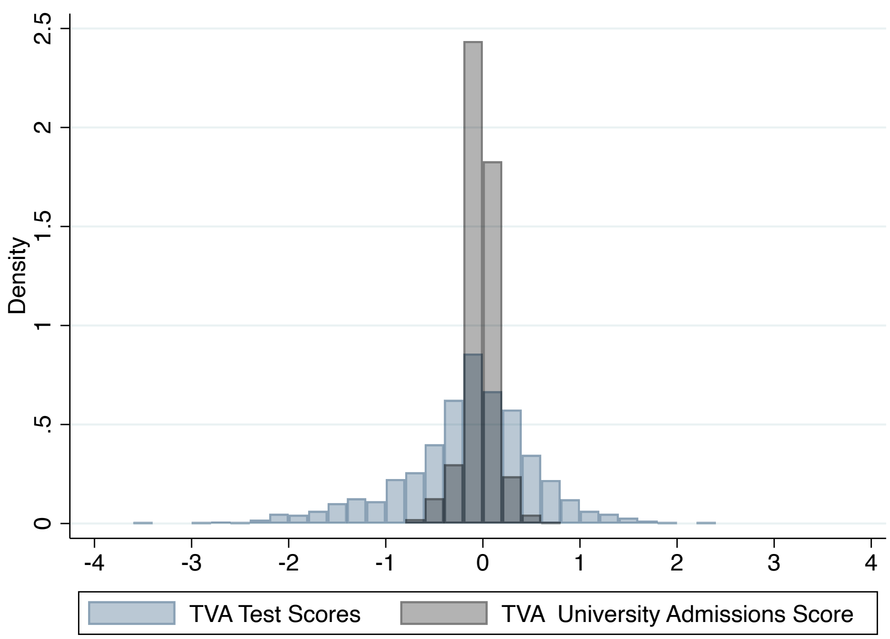

Notes: The top figure shows a binned scatter plot of changes in actual scores and TVA changes. The bottom left figure shows a binned scatter plot of changes in student gender and TVA. The bottom right figure shows a binned scatter plot of changes in student age and TVA. To produce these figures we use $11^{th}$- and $12^{th}$-grade data for the years 2003-2005. These figures pool all available grades and subjects and are constructed using the sample used to estimate the TVA model, which has one observation per student-school-class subject-year cell. The solid line shows the best linear fit estimated on the underlying microdata using OLS.

Figure 4: Histograms of Test-scores VA and Long-Run VAs

Panel A: Overlap of Test-Score TVA and Degree Quality TVA

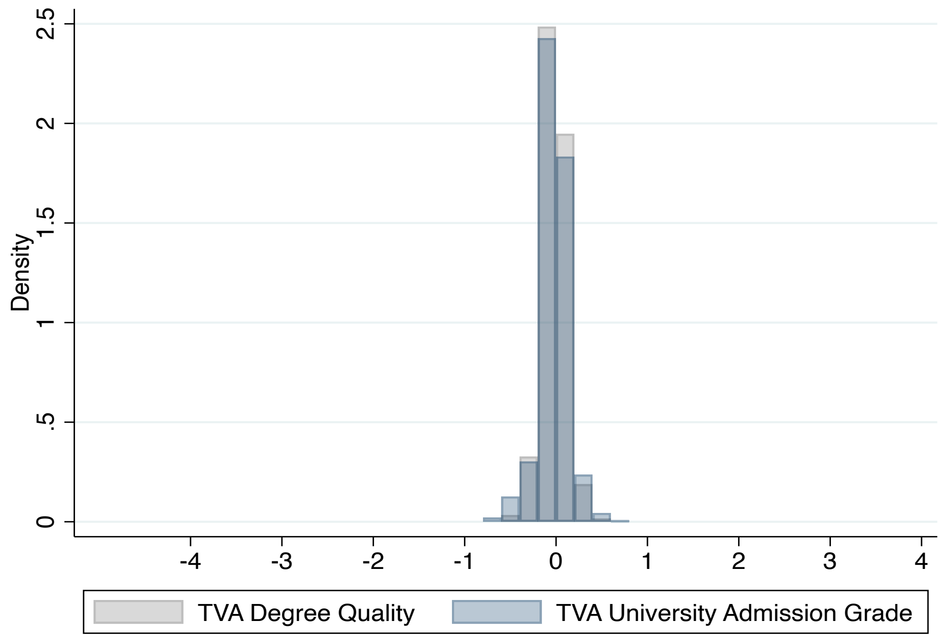



Panel B: Overlap of Test-Score TVA and University Admissions Score TVA



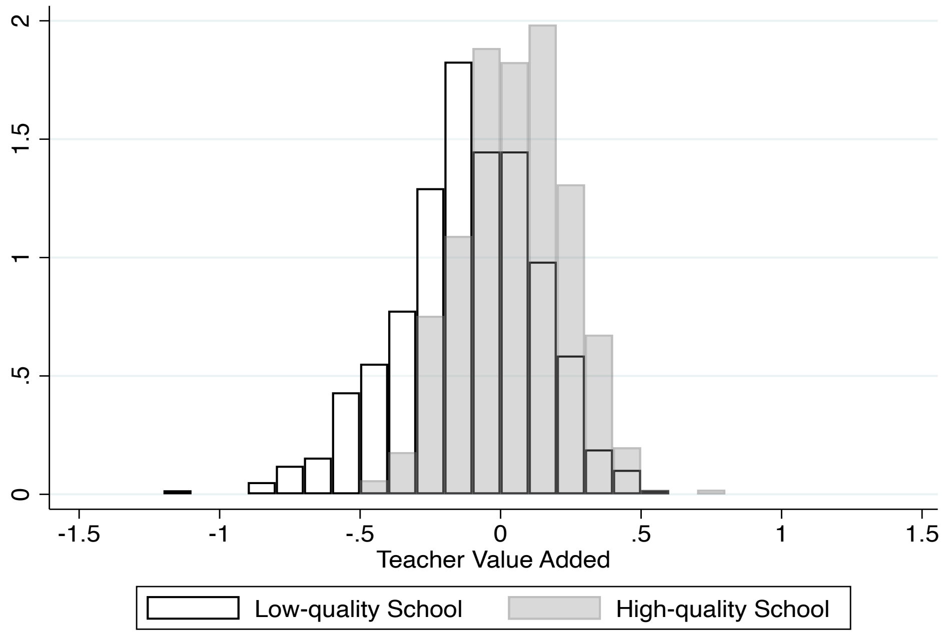

*Notes:* The top figure presents the distribution of test-score TVA when contemporaneous scores are used in the measurement of TVA (*"TVA Test Scores"*) and long-run TVA when students' quality of enrolled degree is used (*"TVA University Degree Quality"*). The bottom figure presents the distribution of test-score TVA when contemporaneous scores are used in the measurement of TVA and long-run TVA when students' university admissions score is used (*"TVA University Admissions Score"*). The distribution of test-score TVA is more spread than that of long-run outcomes TVA. The distributions of all TVA measures are centered around 0.

Figure 5: Histograms of long-term TVA Measures




*Notes:* This figure presents distributions of the two long-term TVA measures. *"TVA University Degree Quality"* is the long-run TVA, which uses students' enrolled degree quality to measure TVA. *"TVA University Admissions Score"* is the long-run TVA, which uses students' university admissions score to measure TVA. There is a significant overlap between the two distributions. Both distributions are centered around 0.

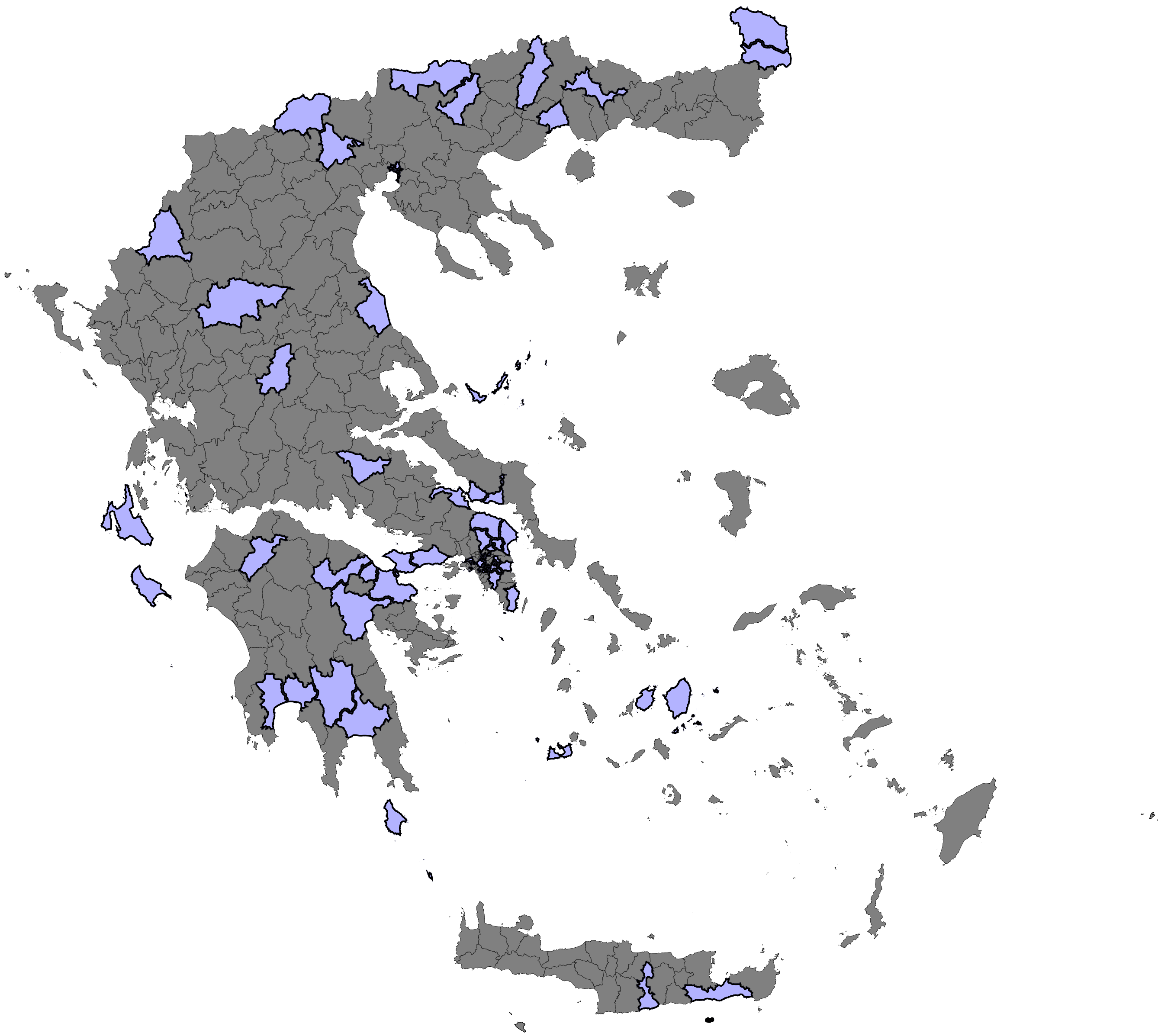
Figure 6:
Teacher Value Added and School Quality




Notes: These two histograms present the distribution of TVA for schools below ("low-quality schools") and above ("high-quality schools") the median school average performance on national university-entrance exams. School quality is determined based on the average school performance of students who attended those high schools in 2003, the first year in the sample.

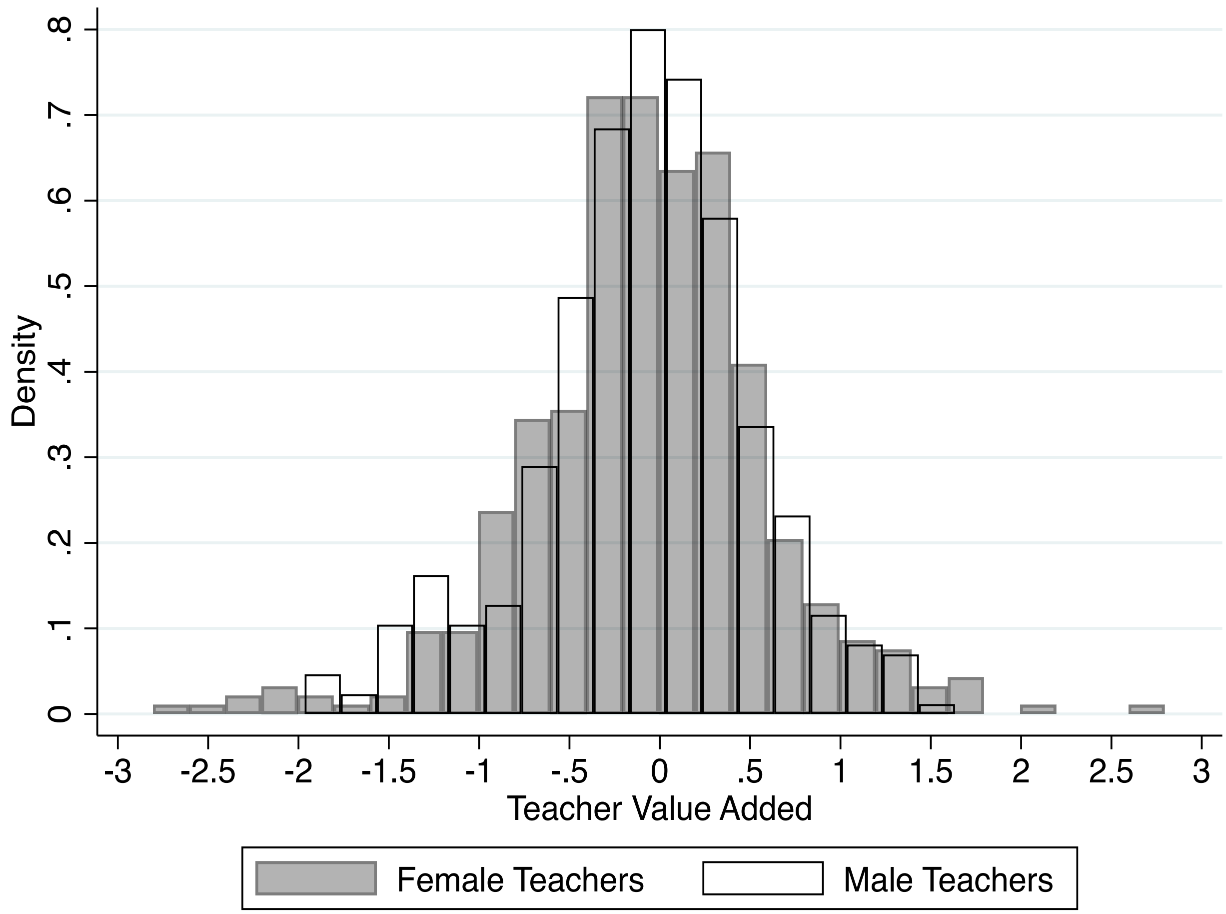
Online Appendix

Figure A1:
MAP OF SCHOOLS IN THE SAMPLE



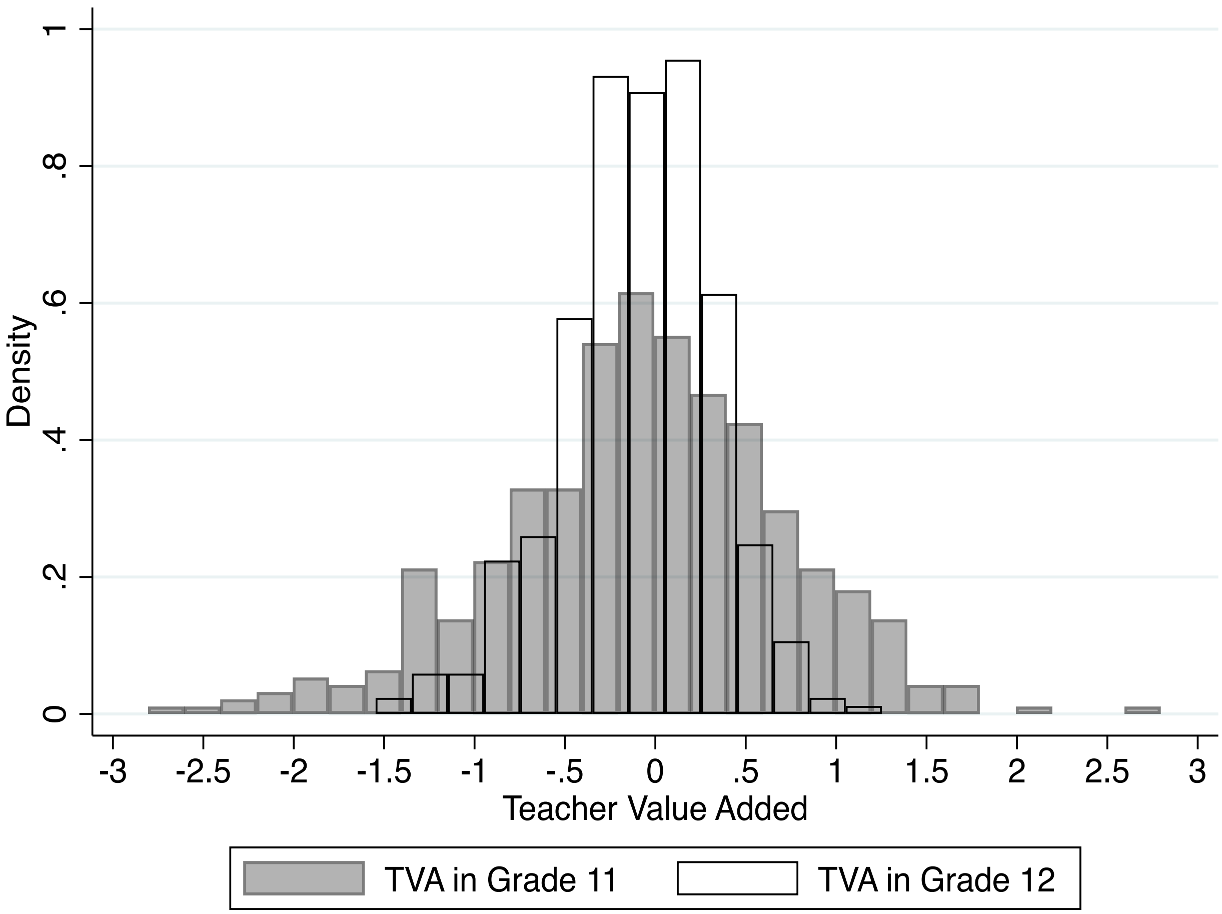
Notes: This figure shows the counties in which high schools in our sample are located.

Figure A2:
TEACHER VALUE ADDED BY TEACHER GENDER



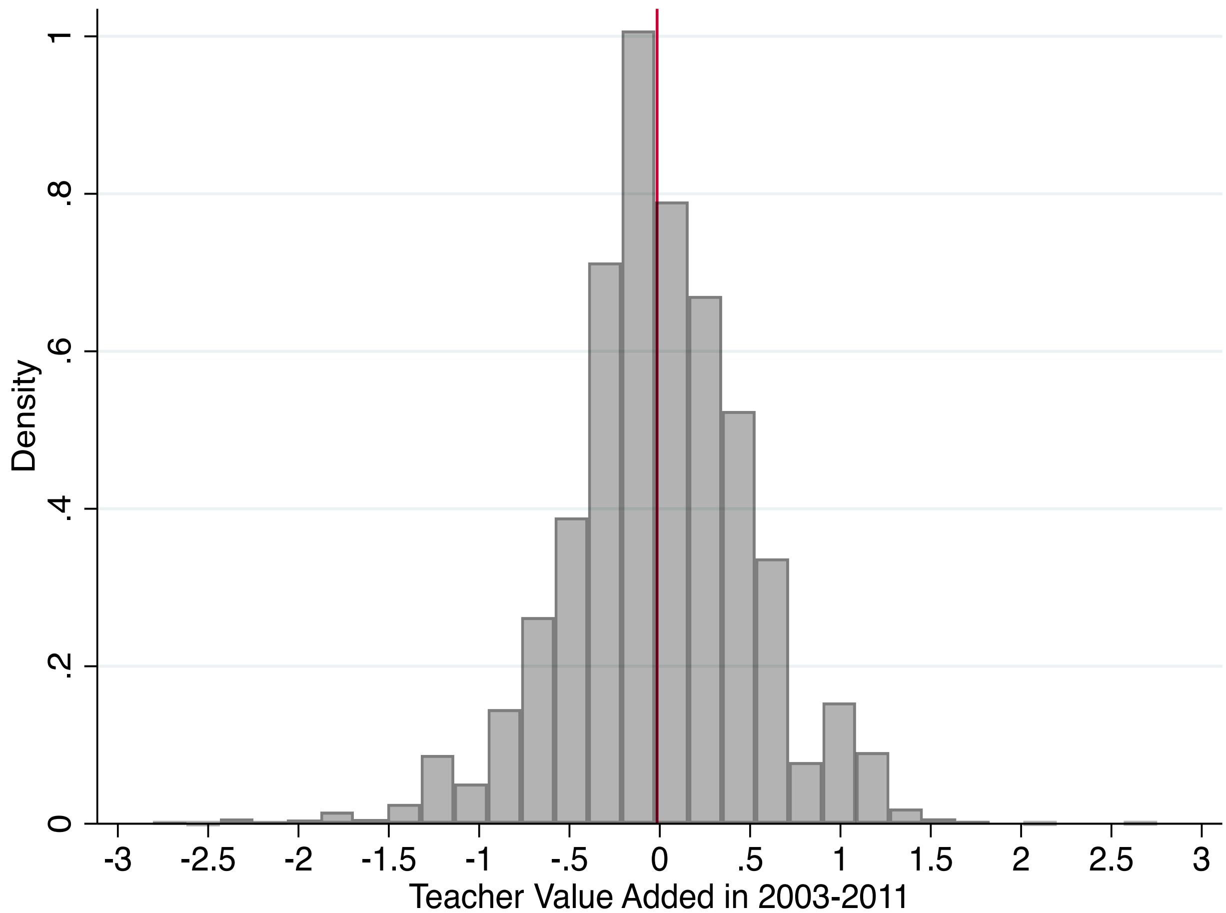

Notes: The histogram shows the distribution of teacher VA for female and male teachers in 2003-2005. Teacher VA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. The mean teacher VA for female teachers is -0.075 (SD=0.668) and for male teachers is -0.095 (SD=0.601).

Figure A3:
Teacher Value Added by Grade



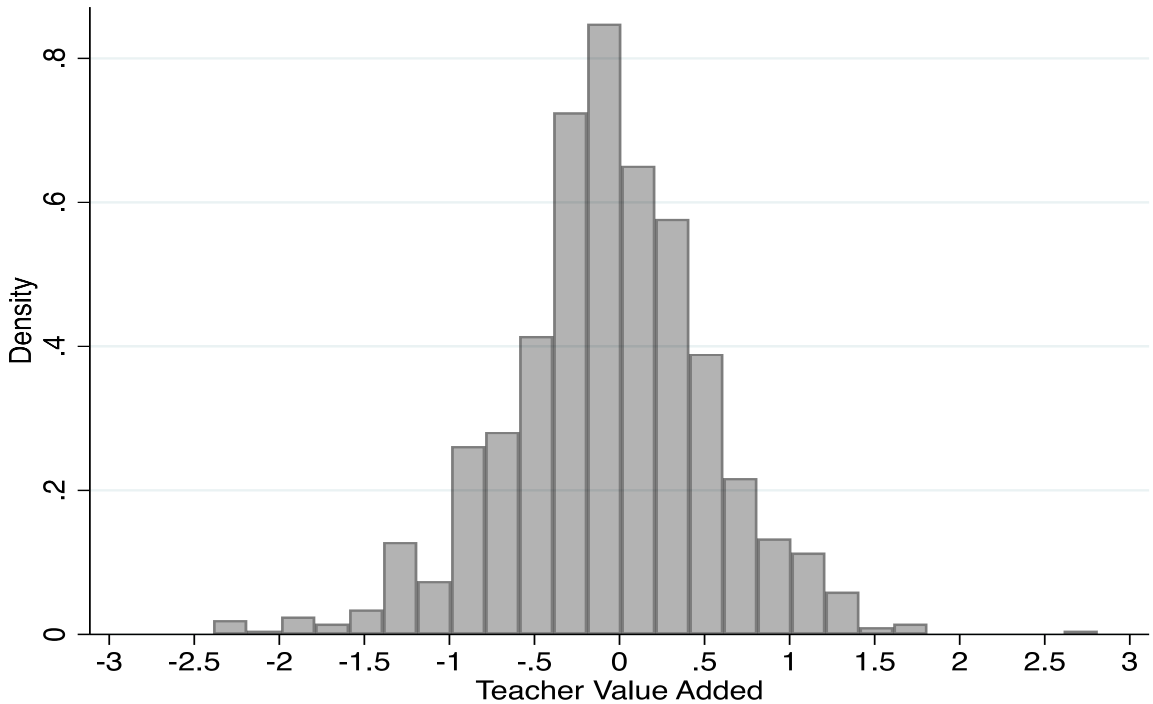

Notes: The histogram shows the distributions of teacher VA for female and male teachers in $11^{th}$ and $12^{th}$ grades. Teacher VA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. The mean teacher VA for $11^{th}$-grade teachers is -0.070 (SD=0.790) and for $12^{th}$-grade teachers is -0.089 (SD=0.042).

Figure A4:
Histogram for Teacher Value Added Measure used in the 2003-2011 Analysis




Notes: This histogram presents the TVA distribution based on test scores. We pool $12^{th}$-grade data for 2003-2011 to derive these value added measures. We use $11^{th}$-grade test scores as the previous-year test scores. The sample includes only students with non-missing previous-year test scores and other requisite controls to estimate the TVA model. Teacher VA is estimated using the baseline control vector, which includes previous-year own-subject test scores; student-level characteristics including age, gender, a binary indicator for being born in the first quarter of the birth year; class size, school enrollment size, and school; and year dummies. The structure of the dataset is one observation per teacher-year combination. We do not display a negative unique outlier value here (or in the histograms that present the 2003-2011 test-score TVA distribution) to maintain a symmetric distribution. However, all values, including the outlier, are included in the analysis.

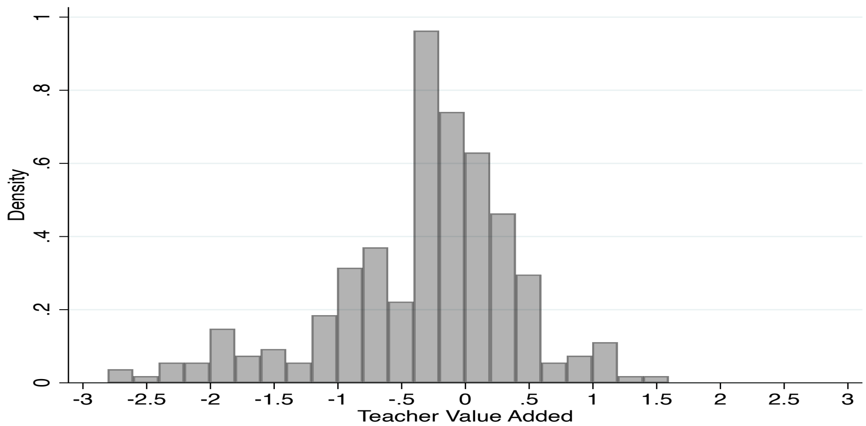
Figure A5:
Histogram of Teacher Value Added in Core Subjects



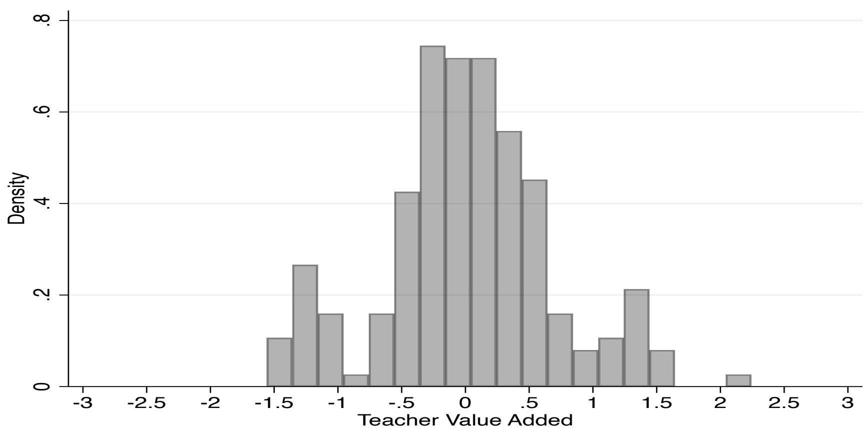

Notes: This histogram shows the teacher VA distribution in the core subjects in $11^{th}$ and $12^{th}$ grades. The core subjects include modern Greek, history, algebra, geometry, and physics in $11^{th}$ grade and modern Greek, history, physics, biology, and mathematics in $12^{th}$ grade. The mean teacher VA for teachers in the core is -0.077 (SD=0.60).

Figure A6:
Histogram of Teacher Value Added by High School Track

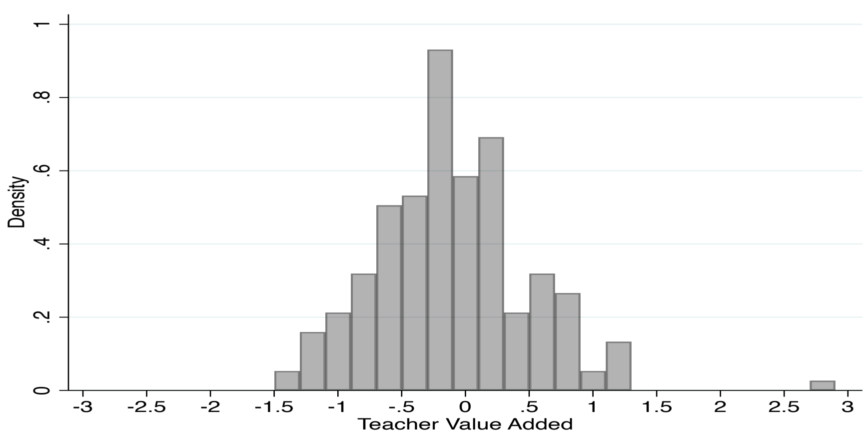
Panel A: Teacher VA in Classics Track




Panel B: Teacher VA in Science Track

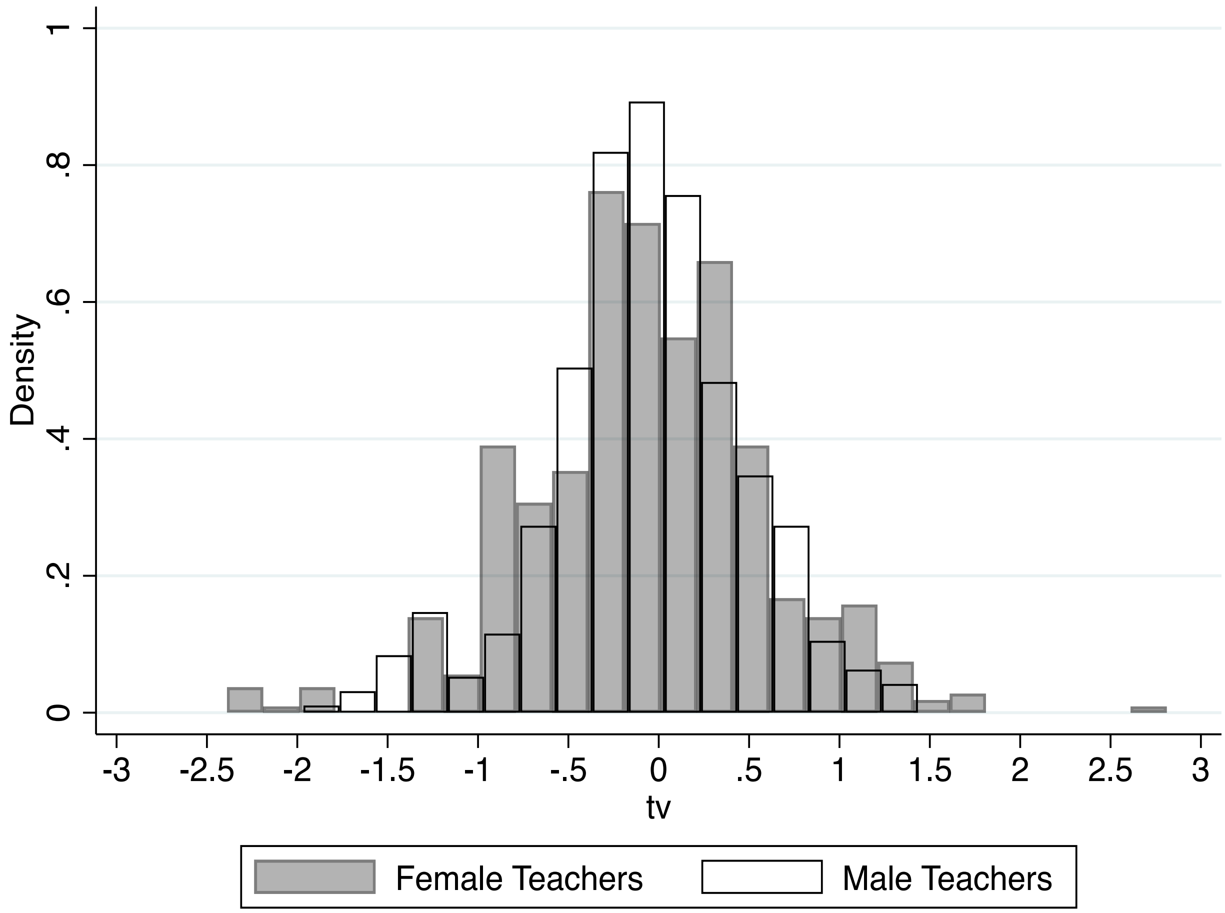



Panel C: Teacher VA in Exact Science Track




Notes: Panels A, B, and C show TVA distributions for the classics, science, and exact science tracks, respectively. Classics track subjects include ancient Greek, philosophy, and Latin in $11^{th}$ grade and ancient Greek, Latin, literature, and history in $12^{th}$ grade. Science track subjects include mathematics, physics, and chemistry in $11^{th}$ grade and biology, mathematics, physics, and chemistry in $12^{th}$ grade. Exact science track subjects include mathematics, physics, and computer science in $11^{th}$ grade and ancient biology, mathematics, physics, business administration, and computer science in $12^{th}$ grade. The mean teacher VA for teachers in the classics track is -0.329 (SD=0.740). The mean teacher VA for teachers in the science track is 0.021 (SD=0.667). The mean teacher VA for teachers in the exact science track is -0.106 (SD=0.608). Data for the period 2003-2005 are used.

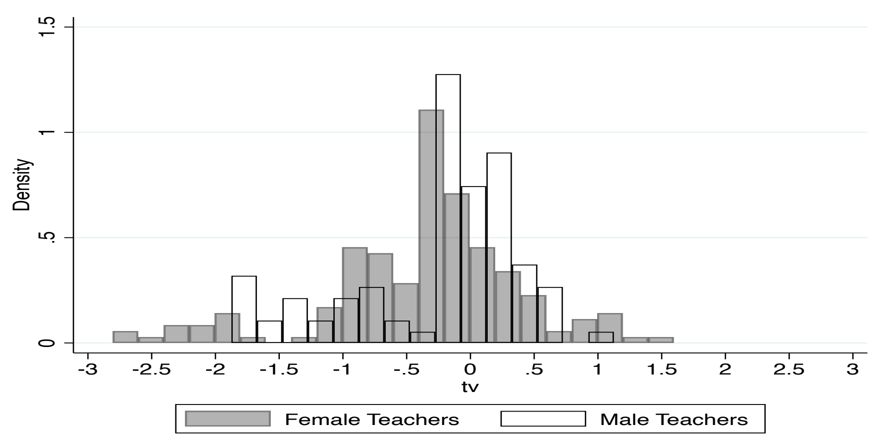
Figure A7:
Histogram of Teacher Value Added by Teacher Gender in Core Subjects



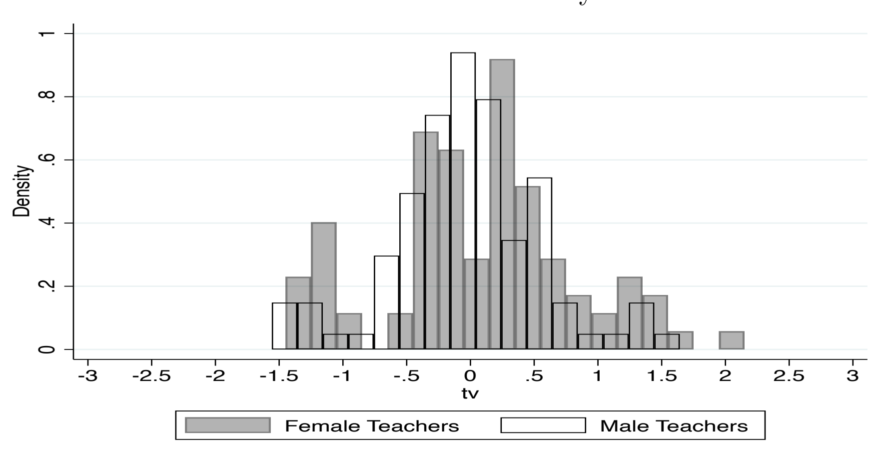

Notes: This histogram shows the distribution of teacher VA in core subjects for female and male teachers separately. The mean teacher VA for female and male teachers in the core subjects is -0.068 (SD=0.63) and -0.086 (SD=0.54), respectively.

Figure A8:
HISTOGRAMS OF TEACHER VALUE ADDED BY HIGH SCHOOL TRACKS AND TEACHER GENDER

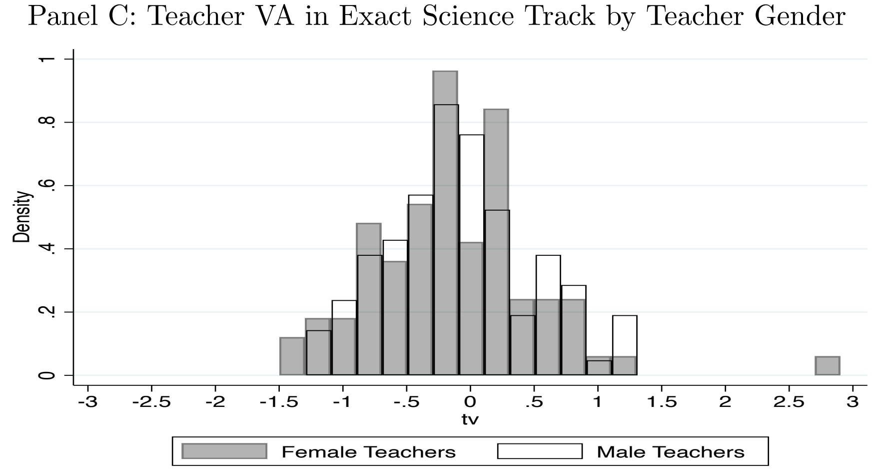
Panel A: Teacher VA in Classics Track by Teacher Gender




Panel B: Teacher VA in Science Track by Teacher Gender




Panel C: Teacher VA in Exact Science Track by Teacher Gender




Notes: These histograms show teacher value added distributions for the three high school tracks by teacher gender. Panels A, B, and C show the TVA distribution by teacher gender for the classics, science, and exact science tracks, respectively. The mean TVA for female and male teachers in the classics track is -0.365 (SD=0.773) and -0.263 (SD=0.675), respectively. The mean TVA for female and male teachers in the science track is 0.077 (SD=0.754) and -0.027 (SD=0.581), respectively. The mean TVA for female and male teachers in the exact science track is -0.134 (SD=0.654) and -0.082 (SD=0.571), respectively. Data for the period 2003-2005 are used

Table A1: Differences Between Study Sample and Population

| | Sample of 21 Schools | | Remaining 1,390 Schools | | | | |
|---|---|---|---|---|---|---|---|
| | Mean | s.d | Mean | s.d | Difference (1)-(3) | s.e | P-value |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | | | | Panel A: 2003-2005 | | | |
| Gender (1=female) | 0.560 | 0.077 | 0.549 | 0.117 | 0.011 | 0.026 | 0.681 |
| High School Graduation Age | 17.872 | 0.131 | 18.198 | 1.505 | -0.326 | 0.328 | 0.321 |
| National Exam Average Test Score (out of 20) | 11.373 | 1.356 | 11.350 | 1.853 | 0.023 | 0.406 | 0.955 |
| ***Track Specialization*** | | | | | | | |
| Classics | 0.385 | 0.102 | 0.399 | 0.160 | -0.015 | 0.035 | 0.673 |
| Science | 0.169 | 0.103 | 0.148 | 0.102 | 0.021 | 0.022 | 0.345 |
| Exact Science | 0.447 | 0.144 | 0.453 | 0.154 | -0.006 | 0.034 | 0.849 |
| Postsecondary Admission Score (out of 20,000) | 13025.811 | 942.347 | 12804.796 | 1431.621 | 221.014 | 313.556 | 0.481 |
| Postsecondary Admission Rate | 0.748 | 0.119 | 0.749 | 0.154 | -0.001 | 0.034 | 0.971 |
| Postcode Net Income (in Euro) | 19113.252 | 4188.484 | 20452.166 | 8361.451 | -1338.913 | 1828.786 | 0.464 |
| Postcode Unemployment Rate | 9.297 | 2.823 | 9.747 | 2.896 | -0.450 | 0.637 | 0.480 |
| Urban | 0.667 | 0.483 | 0.738 | 0.440 | -0.071 | 0.097 | 0.464 |
| | | | | Panel B: 2003-2011 | | | |
| Gender (1=female) | 0.578 | 0.058 | 0.562 | 0.104 | 0.016 | 0.025 | 0.510 |
| High School Graduation Age | 17.913 | 0.081 | 18.419 | 2.186 | -0.506 | 0.515 | 0.326 |
| National Exam Average Test Score (out of 20) | 11.975 | 1.209 | 12.115 | 1.788 | -0.139 | 0.423 | 0.742 |
| ***Track Specialization*** | | | | | | | |
| Classics | 0.398 | 0.069 | 0.420 | 0.141 | -0.021 | 0.033 | 0.523 |
| Science | 0.135 | 0.085 | 0.132 | 0.081 | 0.002 | 0.019 | 0.900 |
| Exact Science | 0.467 | 0.089 | 0.448 | 0.123 | 0.019 | 0.029 | 0.518 |
| Postsecondary Admission Score (out of 20,000) | 13578.768 | 774.927 | 13600.096 | 1370.021 | -21.329 | 323.663 | 0.947 |
| Postsecondary Admission Rate | 0.754 | 0.100 | 0.762 | 0.141 | -0.007 | 0.033 | 0.823 |
| Postcode Net Income (in Euro) | 18845.822 | 4406.896 | 20460.129 | 8267.385 | -1614.307 | 1952.703 | 0.409 |
| Postcode Unemployment Rate | 8.599 | 1.558 | 9.578 | 2.299 | -0.978 | 0.543 | 0.072 |
| Urban | 0.667 | 0.485 | 0.729 | 0.445 | -0.062 | 0.106 | 0.556 |

Notes: Panel A refers to 2003-2005 and Panel B to 2003-2011. Data on all senior high schools (= 1,411) in operation in Greece during the sample period are used. The Hellenic Ministry of Education provides the dataset. *Postcode Net Income* and *Postcode Unemployment Rate* refer to the postcode in which the school is located. Observations are at the school level.

Table A2: Summary Statistics for Sample Used to Estimate Teacher Value Added Models, 2003-2011

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| ***Student Characteristics*** | | | | |
| Gender (1=Female) | 0.567 | 0.495 | 0 | 1 |
| Previous Year Test Scores (std) | -0.059 | 1.009 | -4.383 | 2.713 |
| Age | 17.877 | 0.483 | 17 | 42 |
| No. of Subjects per Student | 9.390 | 2.033 | 1 | 12 |
| ***Tracks of Specialization*** | | | | |
| Classics | 0.362 | 0.481 | 0 | 1 |
| Science | 0.151 | 0.358 | 0 | 1 |
| Exact Science | 0.487 | 0.500 | 0 | 1 |
| ***Class Characteristics*** | | | | |
| Class size | 18.815 | 6.710 | 1 | 37 |
| ***Students Outcomes*** | | | | |
| ***High School*** | | | | |
| Test Score (std) | -0.088 | 1.009 | -3.972 | 2.857 |
| ***University Enrollment*** | | | | |
| University Admissions Score | 12.326 | 4.123 | 2 | 20 |
| Post-secondary Schooling (0/1) | 0.809 | 0.393 | 0 | 1 |
| Academic University Vs Technical School (0/1) | 0.522 | 0.500 | 0 | 1 |
| Post-secondary Degree Quality (Rank 1-100) | 45.689 | 29.075 | 0 | 99.685 |
| Rank of Attending Institution on Degree Preference | 9.675 | 12.385 | 1 | 140 |
| Winning a State Government Scholarship | 0.020 | 0.141 | 0 | 1 |
| ***Teacher Characteristics*** | | | | |
| Teacher VA (2003-2011) | -0.15 | 0.946 | -7.333 | 19.021 |
| Female Teachers | -0.175 | 0.745 | -7.333 | 1.998 |
| Male Teachers | -0.130 | 1.112 | -3.487 | 19.021 |
| Teacher's Gender (1=Female) | 0.501 | 0.500 | 0 | 1 |
| Teacher's Experience (based on Previous Workload) | 7.318 | 5.729 | 2 | 42 |

Notes: All statistics reported are for the sample used to estimate the baseline value added model. Only data for grade 12 are used for the 2003-2011 sample. No. of subjects per student counts the total number of subjects studied in grades 11 and 12. This sample includes only non-missing previous-year test scores and other requisite controls to estimate the TVA model. Student data are from the administrative records of the sample of 21 schools in Greece described in the text. Test scores are standardized z-scores. Age is measured in years and on the day they take the $12^{th}$-grade exam. "Postsecondary Schooling" is a binary indicator that takes the value 1 if a student enrolls in postsecondary schooling and 0 otherwise. The binary indicator "Academic University Vs Technical School" takes the value 1 if the enrolled postsecondary institution is an academic university and 0 if it is a technical school. "Postsecondary Degree Quality 1" is a degree quality based on the university admissions score cutoff. "Postsecondary Degree Quality 2" is a degree's quality based on enrolled students' annual mean national exam performance. "Rank of Attending Institution on Degree List" is the rank of the enrolled option in a student's preference list. The smaller this number is, the more desirable this degree choice for a student is. Teacher characteristics are computed based on the teacher sample. Teacher Value Added estimates are teacher- and year-specific. Student characteristics and outcomes are calculated in a stacked sample with one observation per student-school-class-subject-year cell.

Table A3:
ADDITIONAL SUMMARY STATISTICS FOR THE SAMPLE USED TO ESTIMATE TEACHER VALUE ADDED MODELS, 2003-2011

| | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Schools | 18 | | | | |
| Classes | 310 | | | | |
| Teachers in $12^{th}$ Grade | 292 | | | | |
| Students in $12^{th}$ Grade | 4,828 | | | | |
| Subjects in $12^{th}$ Grade with TVA | | 8.512 | 1.901 | 1 | 12 |
| Teachers per Student in $12^{th}$ Grade with Computed TVA | | 5.627 | 1.433 | 1 | 10 |
| TVA Estimates | 1,019 | | | | |
| Stacked Observations with Computed TVA | 38,259 | | | | |
| Teacher Experience (How Many Classes a Teacher Taught in the Past) | 1,019 | 7.278 | 5.687 | 2 | 42 |
| Female Teachers | 292 | 0.524 | 0.500 | 0 | 1 |
| Male Teachers | 292 | 0.476 | 0.500 | 0 | 1 |

Notes: The dataset is stacked with one observation for each student-school-class-subject-year cell. Teacher VA is calculated with respect to the teacher, year, and grade cell. Teacher VA is calculated for teachers who teach the same grade for at least 2 years. If a teacher teaches for only 1 year or one grade we cannot estimate their teacher VA. Teacher experience measures the previous workload in the study period. It calculates how many times (class-year-subject cell) a teacher has taught in the corresponding period. A teacher sample is used to compute the share of female teachers.

Table A4: Additional Balancing Test of Teacher Characteristics on Students' Pre-Assignment Characteristics, 2003-2005

| | Teacher Characteristics | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Gender (=1 Female) | Value Added $_t$ | Value Added$_{t\text{-}1}$ | Experience |
| | (1) | (2) | (3) | (4) |
| ***Student Characteristics*** | | | | |
| GPA in $10^{th}$ Grade | 0.002 | -0.001 | -0.000 | 0.014 |
| | (0.003) | (0.002) | (0.002) | (0.031) |
| Mathematics in $10^{th}$ Grade | 0.000 | 0.003 | 0.002 | 0.006 |
| | (0.002) | (0.002)* | (0.002) | (0.031) |
| English in $10^{th}$ Grade | -0.001 | -0.000 | -0.000 | 0.001 |
| | (0.001) | (0.001) | (0.001) | (0.016) |
| Gender (=1 if Female) | -0.002 | -0.000 | 0.001 | 0.008 |
| | (0.002) | (0.002) | (0.002) | (0.043) |
| Age | 0.000 | -0.000 | 0.001 | 0.026 |
| | (0.001) | (0.001) | (0.002) | (0.022) |
| *N* | 36,323 | 36,323 | 29,310 | 36,323 |
| F-test for Joint Significance | 0.44 | 0.87 | 0.45 | 0.66 |
| P-value for F-test | 0.819 | 0.510 | 0.808 | 0.658 |
| Grade FE | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ |

Notes: The same regression generates all estimated coefficients in each column. Thus, estimates in this table come from four separate regressions, where the outcome variables are reported in each column headings. The table reports OLS coefficients from separate regressions of each teacher's characteristics on all student pre-assignment characteristics. Students' pre-assignment characteristics include gender (=1 if female), previous-year test scores and age. Teacher characteristics include teacher gender (=1 if female), teacher quality measured in the same year (proxied by a teacher's value added in year t), teacher quality measured in the previous year (proxied by a teacher's value added in year t-1), and teaching experience based on the previous workload. All regressions include controls for grade fixed effects, year fixed effects, track fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Balancing Test of Students' Pre-Assignment Characteristics on Teacher Characteristics, Track Classes Only, 2003-2005

| | Student Characteristics | | | | |
|---|---|---|---|---|---|
| | GPA in $10^{th}$ Grade (1) | Mathematics in $10^{th}$ Grade (2) | English in $10^{th}$ Grade (3) | Gender (=1 Female) (4) | Age (5) |
| ***Teacher Characteristics*** | | | | | |
| Gender (=1 Female) | 0.026 | 0.024 | 0.014 | 0.003 | 0.000 |
| | (0.017) | (0.015) | (0.014) | (0.007) | (0.006) |
| *N* | 12,058 | 12,058 | 10,814 | 12,909 | 12,909 |
| Value Added$_{\mathrm{t}}$ | 0.014 | 0.015 | 0.014 | 0.007 | 0.001 |
| | (0.023) | (0.021) | (0.016) | (0.009) | (0.009) |
| *N* | 12,058 | 12,058 | 10,814 | 12,909 | 12,909 |
| Value Added$_{\mathrm{t-1}}$ | 0.031 | 0.019 | 0.014 | 0.013 | 0.005 |
| | (0.021) | (0.019) | (0.016) | (0.009) | (0.015) |
| *N* | 9,589 | 9,589 | 8,599 | 10,317 | 10,317 |
| Experience | 0.002 | 0.001 | 0.000 | -0.000 | -0.000 |
| | (0.001)* | (0.001)* | (0.001) | (0.001) | (0.000) |
| *N* | 12,058 | 12,058 | 10,814 | 12,909 | 12,909 |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each estimate in this table is generated from a different regression. The table reports OLS coefficients from separate regressions of each student's pre-assignment characteristics on each of the teacher's characteristics. The scores in (GPA, mathematics, and English) are standardized and have a zero mean and a standard deviation of one. We use the average standardized performance in algebra and geometry in $10^{th}$ grade for mathematics. Students' pre-assigned characteristics include gender (=1 if female), previous-year test scores, and age. The dependent variable is the GPA in grade 10 in column (1), test scores in mathematics in grade 10 in column (2), test scores in English in grade 10 in column (3), a binary indicator for the gender of the student (=1 if female) in column (4), and age in column (5). The independent variables are listed vertically and include the respective teacher characteristics. In particular, we use teacher gender (=1 if female), assigned teacher's previous-year quality (measured by the assigned teacher's previous-year value added), and teaching experience based on the previous workload (measured by how many times the assigned teacher has taught in the sample period of 2003-2005). All regressions condition on track fixed effects, year fixed effects, grade fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A6: Additional Balancing Test of Teacher Characteristics on Students' Pre-Assignment Characteristics, Track Classes Only, 2003-2005

| | Teacher Characteristics | | | |
|---|---|---|---|---|
| | Gender (=1 Female) | Value Added$_t$ | Value Added$_{t\text{-}1}$ | Experience |
| ***Student Characteristics*** | | | | |
| GPA in $10^{th}$ Grade | 0.003 | -0.003 | 0.005 | 0.049 |
| | (0.005) | (0.005) | (0.004) | (0.077) |
| Mathematics in $10^{th}$ Grade | 0.006 | 0.006 | -0.002 | 0.039 |
| | (0.005) | (0.005) | (0.004) | (0.073) |
| English in $10^{th}$ Grade | -0.001 | 0.001 | -0.001 | -0.023 |
| | (0.002) | (0.002) | (0.002) | (0.033) |
| Gender (=1 if Female) | 0.002 | 0.002 | 0.004 | -0.079 |
| | (0.006) | (0.004) | (0.004) | (0.114) |
| Age | 0.001 | 0.001 | 0.003 | 0.014 |
| | (0.004) | (0.004) | (0.006) | (0.071) |
| *N* | 10,814 | 10,814 | 8,599 | 10,814 |
| F-test for Joint Significance | 1.66 | 0.41 | 0.55 | 0.48 |
| P-value of F-test | 0.162 | 0.840 | 0.739 | 0.789 |
| Grade FE | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ |

Notes: The same regression generates all estimated coefficients in each column. Estimates in this table come from four separate regressions, where the outcome variables are reported in each column. The table reports OLS coefficients from separate regressions of each teacher's characteristics on all students' pre-assignment characteristics. Students' pre-assignment characteristics include gender (=1 if female), previous-year test scores, and age. Teacher characteristics include teacher gender (=1 if female), teacher quality measured in the same year (proxied by a teacher's value added in year t), teacher quality measured in the previous year (proxied by a teacher's value added in year t-1), and teaching experience based on previous workload (measured by how many times a teacher teaches in the sample period of 2003-2005). All regressions condition on year fixed effects, grade fixed effects, track fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7: Balancing Test of Students' Pre-Assignment Characteristics on Teacher Characteristics, Core Classes Only, 2003-2005

| | Student Characteristics | | | | |
|---|---|---|---|---|---|
| | GPA in $10^{th}$ Grade (1) | Mathematics in $10^{th}$ Grade (2) | English in $10^{th}$ Grade (3) | Gender (=1 Female) (4) | Age (5) |
| ***Teacher Characteristics*** | | | | | |
| Gender (=1 Female) | 0.006 | 0.004 | 0.002 | -0.003 | -0.001 |
| | (0.007) | (0.006) | (0.006) | (0.002) | (0.002) |
| *N* | 28,490 | 28,490 | 25,509 | 29,823 | 29,823 |
| Value Added$_{t}$ | 0.005 | 0.006 | 0.001 | 0.000 | -0.004 |
| | (0.008) | (0.007) | (0.008) | (0.003) | (0.003) |
| *N* | 28,490 | 28,490 | 25,509 | 29,823 | 29,823 |
| Value Added$_{t-1}$ | 0.011 | 0.010 | 0.006 | 0.001 | -0.002 |
| | (0.007) | (0.007) | (0.006) | (0.004) | (0.005) |
| *N* | 22,908 | 22,908 | 20,518 | 24,040 | 24,040 |
| Experience | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| *N* | 28,490 | 28,490 | 25,509 | 29,823 | 29,823 |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each estimate in this table is generated from a different regression. The table reports OLS coefficients from separate regressions of each student's pre-assignment characteristics on each of the teacher's characteristics. The scores in $10^{th}$ grade (GPA, mathematics and English) are standardized and have a zero mean and a standard deviation of 1. We use the average standardized performance in algebra and geometry in $10^{th}$ grade for mathematics. The subject of Economics (optional) is also included in this table. Students pre-assigned characteristics include gender (=1 if female), previous-year test scores, and age. The dependent variable is the GPA in grade 10 in column (1), test scores in mathematics in grade 10 in column (2), test scores in English in grade 10 in column (3), a binary indicator for the gender of the student (=1 if female) in column (4), and age in column (5). The independent variables are listed vertically and include the respective teacher characteristics. In particular, we use teacher gender (=1 if female), assigned teacher's previous-year quality (measured by the assigned teacher's previous-year value added), and teaching experience based on the previous workload (measured by how many times the assigned teacher has taught in the sample period of 2003-2005). All regressions condition on track fixed effects, year fixed effects, grade fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A8: ADDITIONAL BALANCING TEST OF TEACHER CHARACTERISTICS ON STUDENTS' PRE-ASSIGNMENT CHARACTERISTICS, CORE CLASSES ONLY, 2003-2005

| | Teacher Characteristics | | | |
|---|---|---|---|---|
| | Gender (=1 Female) (1) | Value Added$_{t}$ (2) | Value Added$_{t-1}$ (3) | Experience (4) |
| ***Student Characteristics*** | | | | |
| GPA in $10^{th}$ Grade | 0.002 | -0.001 | -0.000 | 0.014 |
| | (0.003) | (0.002) | (0.002) | (0.031) |
| Mathematics in $10^{th}$ Grade | 0.000 | 0.003 | 0.002 | 0.006 |
| | (0.002) | (0.002)* | (0.002) | (0.031) |
| English in $10^{th}$ Grade | -0.001 | -0.000 | -0.000 | 0.001 |
| | (0.001) | (0.001) | (0.001) | (0.016) |
| Gender (=1 if Female) | -0.002 | -0.000 | 0.001 | 0.008 |
| | (0.002) | (0.002) | (0.002) | (0.043) |
| Age | 0.000 | -0.000 | 0.001 | 0.026 |
| | (0.001) | (0.001) | (0.002) | (0.022) |
| $N$ | 25,509 | 25,509 | 20,518 | 25,509 |
| F-test for Joint Significance | 1.45 | 1.18 | 0.81 | 1.56 |
| P-value of F-test | 0.220 | 0.880 | 0.330 | 0.188 |
| Grade FE | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ |

Notes: The same regression generates all estimates in each column. So estimates in this table come from four separate regressions, where outcome variables are reported in each column. The table reports OLS coefficients from separate regressions of each teacher's characteristics on all student pre-assignment characteristics. The subject of Economics (optional) is also included in this table. Students' pre-assignment characteristics include: gender (=1 if female), previous-year test scores, and age. Teacher characteristics include teacher gender (=1 if female), teacher quality measured in the same year (proxied by a teacher's value added in year t), teacher quality measured in the previous year (proxied by a teacher's value added in year t-1), and teaching experience based on the previous workload (measured by how many times a teacher teaches in the sample period of 2003-2005). All regressions condition on year fixed effects, grade fixed effects, track fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9: Balancing Test of Students' Pre-Assignment Characteristics on Teacher Characteristics, All Classes, 2003-2011

| | Student Characteristics | | | | |
|---|---|---|---|---|---|
| | GPA in $10^{th}$ Grade (1) | Mathematics in $10^{th}$ Grade (2) | English in $10^{th}$ Grade (3) | Gender (=1 Female) (4) | Age (5) |
| ***Teacher Characteristics*** | | | | | |
| Gender (=1 Female) | 0.014 | 0.042 | 0.038 | -0.008 | -0.002 |
| | (0.009) | (0.010)*** | (0.016)** | (0.005)* | (0.005) |
| *N* | 30,753 | 13,720 | 11,597 | 33,219 | 33,219 |
| Value $\text{Added}_{\text{t}}$ | 0.010 | 0.014 | 0.001 | -0.007 | 0.008 |
| | (0.008) | (0.008)* | (0.007) | (0.005) | (0.006) |
| *N* | 12,058 | 13,720 | 11,597 | 33,219 | 33,219 |
| Value $\text{Added}_{\text{t-1}}$ | 0.005 | 0.006 | 0.005 | -0.002 | -0.007 |
| | (0.007) | (0.008) | (0.012) | (0.003) | (0.005) |
| *N* | 26,370 | 12,661 | 10,765 | 28,502 | 28,502 |
| Experience | -0.002 | -0.003 | -0.003 | 0.000 | 0.000 |
| | (0.001)** | (0.001)*** | (0.002) | (0.000) | (0.000) |
| *N* | 30,753 | 13,720 | 11,597 | 33,219 | 33,219 |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each estimate in this table is generated from a different regression. The table reports OLS coefficients from separate regressions of each student pre-assignment characteristic on each of the teacher's characteristics. The scores in $10^{th}$ grade (GPA, Mathematics, and English) are standardized and have a zero mean and a standard deviation of 1. We use the average standardized performance in algebra and geometry in $10^{th}$ grade for mathematics. Students' pre-assigned characteristics include gender (=1 if female), previous-year test scores, and age. The dependent variable is the respective teachers' characteristics, including teacher gender (1=female), assigned teacher's previous-year quality (measured by the assigned teacher's previous-year value added), and teaching experience based on the previous workload. All regressions condition on track fixed effects, year fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10: Additional Balancing Test of Teacher Characteristics on Students' Pre-Assignment Characteristics, All Classes, 2003-2011

| | Teacher Characteristics | | | |
|---|---|---|---|---|
| | Gender (=1 Female) (1) | Value Added$_t$ (2) | Value Added$_{t-1}$ (3) | Experience (4) |
| ***Student Characteristics*** | | | | |
| GPA in $10^{th}$ Grade | 0.003 | 0.030 | 0.018 | -0.020 |
| | (0.005) | (0.008)*** | (0.009)** | (0.064) |
| Mathematics in $10^{th}$ Grade | 0.008 | -0.013 | -0.011 | -0.025 |
| | (0.005) | (0.007)* | (0.008) | (0.044) |
| English in $10^{th}$ Grade | 0.002 | -0.012 | -0.004 | -0.014 |
| | (0.004) | (0.004)*** | (0.006) | (0.035) |
| Gender (=1 if Female) | 0.006 | -0.016 | -0.003 | -0.079 |
| | (0.006) | (0.008)* | (0.007) | (0.047)* |
| Age | -0.002 | -0.001 | -0.011 | -0.000 |
| | (0.008) | (0.020) | (0.011) | (0.050) |
| *N* | 11,597 | 11,597 | 10,765 | 11,597 |
| F-test for Joint Significance | 2.64 | 2.42 | 1.39 | 1.83 |
| P-value of F-test | 0.032 | 0.046 | 0.240 | 0.120 |
| Grade FE | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ |

Notes: The same regression generates all estimated coefficients in each column. Thus estimates in this table come from four different regressions, where outcome variables are reported in each column. The table reports OLS coefficients from separate regressions of each teacher's characteristics on all student pre-assignment characteristics. Students' pre-assignment characteristics include: gender (=1 if female), previous-year test scores, and age. Teacher characteristics include teacher gender (=1 if female), teacher quality measured in the same year (proxied by TVA in year t), teacher quality measured in the previous year (proxied by a teacher's value added in year t-1), and teaching experience based on the previous workload (measured by how many times a teacher taught in the sample period). All regressions condition on track fixed effects, year fixed effects, and class fixed effects. Robust standard errors clustered at school and cohort levels in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Number of Available Classroom Slots and Number of Teachers to be Assigned to Classrooms per Speciality in Each School, Grade, and Year

Panel A: Classroom Slots per Teaching Specialty in each School-Grade-Year

| | Mean | SD |
|---|---|---|
| Classroom Slots in Mathematics | 10.153 | 3.752 |
| Classroom Slots Physics, Chemistry and Biology | 10.815 | 4.020 |
| Classroom Slots Language and History | 9.971 | 3.740 |

Panel B: Teachers per Teaching Specialty in each School-Grade-Year

| | Mean | SD |
|---|---|---|
| Teachers in Mathematics | 2.381 | 1.043 |
| Teachers in Physics, Chemistry and Biology | 3.757 | 1.595 |
| Teachers in Language and History | 3.754 | 1.554 |

Notes: Panel A shows the mean and standard deviation of the available classroom slots in each teaching specialty within schools, grades, and years in the study sample. These statistics describe the classroom slots principals must fill with teachers of specific subjects. There are three main teaching specialties in the Greek education system: (a) mathematics (teaching categorization ΠΕ03); (b) physics, chemistry and biology (teaching categorization ΠΕ04); and (c) language and history (teaching categorization ΠΕ02). Teachers specializing in mathematics are trained to teach all mathematics courses in the core and tracks. Teachers specializing in physics, chemistry, and biology are trained to teach all courses related to physics, chemistry, and biology in the core and the tracks. Teachers specializing in language and history are trained to teach all language subjects (modern Greek, ancient Greek, Latin) and history courses in the core and track. Classroom slots refer to core and track subjects. In particular, classroom slots in mathematics refer to (core and track) classrooms that need to be assigned teachers in the following subjects: (i) in grade 11: algebra and geometry in the core and mathematics in the science and exact science tracks and (ii) in grade 12: mathematics in the core and mathematics in the science and exact science tracks. Classroom slots in physics, chemistry, and biology refer to (core and track) classrooms which need to be assigned teachers in the following specific subjects: i) in grade 11: physics in the core, mathematics, physics and chemistry in the science track, and mathematics and physics in the exact science track and ii) in grade 12: physics and biology in the core, biology, mathematics, physics, and chemistry in the science track, and mathematics and physics in the exact science track. Classroom slots in language and history refer to (core and track) classrooms that need to be assigned teachers in the following specific subjects: (i) in grade 11: modern Greek, ancient Greek and history in the core, ancient Greek, philosophy, and Latin in the classics track; and (ii) in grade 12: modern Greek and history in the core and modern Greek, ancient Greek, philosophy and history in the classics track. Panel B shows the mean and standard deviation of teachers' availability per teaching specialty within schools, grades, and years. Teachers are assigned to schools. Principals and the school board have to assign teachers to classrooms based on a computerized algorithm. These statistics describe the availability of teachers in each teaching specialization whom principals allocate to available classroom slots per school-year-grade. Data from the 2003-2005 sample are used in this table.

Table A12: Main Estimates on High School Outcomes Accounting for Empirical Bayes Correction and Two-Step Bootstrapped Standard Errors

| | Subject Specific National Score $11^{th}$ Grade 2003-2005 | Subject Specific National Score $12^{th}$ Grade 2003-2005 | Subject Specific National Score, Stacked $11^{th} + 12^{th}$ Grade 2003-2005 | Subject Specific National Score $12^{th}$ Grade 2003-2011 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Teacher VA | 0.215 | 0.204 | 0.219 | 0.244 |
| | (0.015)*** | (0.022)*** | (0.012)*** | (0.025)*** |
| $N$ | 23,566 | 23,566 | 42,731 | 38,244 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Grade FE | | | ✓ | |
| Baseline FE (Year, Track, Subject, Class, Student FE) | ✓ | ✓ | ✓ | ✓ |

Notes: We replicate the main findings of the effects of TVA on high school outcomes while accounting for two correction methods. First, we use an empirical Bayes (EB) shrinkage estimation approach to address potential sampling error because, for some teachers, TVA estimates are based on small samples. Second, we use a two-step bootstrapping method to correct for the fact that the main variable of interest (Teacher VA) is a generated regressor. Details of both correction methods are described in the text.

Table A13: Robustness of Main Effects: The Effect of Teacher Value Added on High School Outcomes, Controls Added Gradually to the Benchmark Specification

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Subject Specific Score, $11^{th}$ and $12^{th}$ (Stacked), 2003-2005 | 0.239 | 0.239 | 0.229 | 0.205 |
| | (0.024)*** | (0.024)*** | (0.025)*** | (0.021)*** |
| $N$ | 42,734 | 42,734 | 42,732 | 42,731 |
| Baseline Controls FE (Year, Subject, Grade, School, Track) | ✓ | ✓ | ✓ | ✓ |
| Student Characteristics & Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Teacher Characteristics | | ✓ | ✓ | ✓ |
| Class FE | | | ✓ | ✓ |
| Student FE | | | | ✓ |
| [1em] Subject Specific Score in 12th, 2003-2011 | 0.196 | 0.195 | 0.206 | 0.192 |
| | (0.037)*** | (0.037)*** | (0.033)*** | (0.040)*** |
| $N$ | 38,258 | 38,258 | 38,258 | 38,244 |
| Baseline Controls FE (Year, Subject, School, Track) | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Student Characteristics & Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Teacher Characteristics | | ✓ | ✓ | ✓ |
| Class FE | | | ✓ | ✓ |
| Student FE | | | | ✓ |

Notes: Each row reports coefficients from separate OLS regressions of the outcome variable, which is reported on the left, on the TVA measure, while we gradually add controls to the benchmark specification as reported in the table. Standard errors clustered by school and cohort levels are in parentheses. TVA is estimated using the baseline control vector described in the text. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A14: Robustness of Main Results: The Effect of Teacher Value Added on Longer-Term Outcomes, Controls Added Gradually to the Benchmark Specification

| | | | | |
|---|---|---|---|---|
| | Panel A: Average TVA Based on Test Scores | | | |
| University Admissions Score | 1.720<br>(0.190)*** | 1.728<br>(0.189)*** | 1.827<br>(0.172)*** | 1.762<br>(0.172)*** |
| Postsecondary Degree Quality | 10.660<br>(1.282)*** | 10.694<br>(1.292)*** | 14.454<br>(1.428)*** | 14.171<br>(1.419)*** |
| Rank of Attending Institution on Degree Preference List | 2.223<br>(0.449)*** | 2.231<br>(0.449)*** | 2.879<br>(0.482)*** | 2.779<br>(0.477)*** |
| Enrollment in Postsecondary Schooling (0/1) | 0.114<br>(0.017)*** | 0.114<br>(0.017)*** | 0.120<br>(0.016)*** | 0.113<br>(0.016)*** |
| Academic University Vs Technical School (0/1) | 0.094<br>(0.021)*** | 0.095<br>(0.021)*** | 0.141<br>(0.023)*** | 0.142<br>(0.024)*** |
| Winning State Government Scholarship (0/1) | 0.031<br>(0.005)*** | 0.031<br>(0.005)*** | 0.032<br>(0.005)*** | 0.033<br>(0.005)*** |
| *N* | 4,494 | 4,494 | 4,494 | 4,494 |
| Baseline Controls FE (Year, School Track) | ✓ | ✓ | ✓ | ✓ |
| Student Characteristics | | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | | | ✓ | ✓ |
| Teacher Characteristics | | | | ✓ |
| | Panel B: Average TVA Based on Degree Quality | | | |
| University Admissions Score | 5.278<br>(0.932)*** | 5.417<br>(0.950)*** | 5.683<br>(0.793)*** | 5.496<br>(0.770)*** |
| Postsecondary Degree Quality | 26.768<br>(5.898)*** | 27.183<br>(5.917)*** | 38.112<br>(5.479)*** | 37.514<br>(5.452)*** |
| Rank of Attending Institution on Degree Preference List | 2.503<br>(2.484) | 2.530<br>(2.452) | 4.179<br>(2.277)* | 3.990<br>(2.226)* |
| Enrollment in Postsecondary Schooling (0/1) | 0.393<br>(0.080)*** | 0.403<br>(0.080)*** | 0.419<br>(0.075)*** | 0.406<br>(0.074)*** |
| Academic University Vs Technical School (0/1) | 0.252<br>(0.112)** | 0.254<br>(0.112)** | 0.397<br>(0.111)*** | 0.393<br>(0.110)*** |
| Winning State Government Scholarship (0/1) | 0.125<br>(0.045)*** | 0.125<br>(0.045)*** | 0.131<br>(0.043)*** | 0.127<br>(0.045)*** |
| *N* | 4,494 | 4,494 | 4,494 | 4,494 |
| Baseline Controls FE (Year, School Track) | ✓ | ✓ | ✓ | ✓ |
| Student Characteristics | | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | | | ✓ | ✓ |
| Teacher Characteristics | | | | ✓ |
| | Panel C: Average TVA Based on University Admissions Score | | | |
| University Admissions Score | 3.956<br>(0.679)*** | 4.030<br>(0.687)*** | 4.455<br>(0.576)*** | 4.200<br>(0.563)*** |
| Postsecondary Degree Quality | 23.145<br>(5.066)*** | 23.292<br>(5.060)*** | 31.825<br>(4.396)*** | 30.768<br>(4.417)*** |
| Rank of Attending Institution on Degree Preference List | 3.264<br>(2.016) | 3.210<br>(2.003) | 4.501<br>(1.813)** | 4.013<br>(1.764)** |
| Enrollment in Postsecondary Schooling (0/1) | 0.290<br>(0.062)*** | 0.295<br>(0.061)*** | 0.320<br>(0.059)*** | 0.301<br>(0.057)*** |
| Academic University Vs Technical School (0/1) | 0.257<br>(0.092)*** | 0.257<br>(0.091)*** | 0.369<br>(0.090)*** | 0.359<br>(0.089)*** |
| Winning State Government Scholarship (0/1) | 0.100<br>(0.034)*** | 0.099<br>(0.034)*** | 0.107<br>(0.033)*** | 0.103<br>(0.035)*** |
| *N* | 4,494 | 4,494 | 4,494 | 4,494 |
| Baseline Controls FE (Year, School Track) | ✓ | ✓ | ✓ | ✓ |
| Student Characteristics | | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | | | ✓ | ✓ |
| Teacher Characteristics | | | | ✓ |

Notes: Each row reports coefficients from separate OLS regressions of the outcome variable, which is reported on the left, on the TVA measure, while we gradually add controls to the benchmark specification as reported in the table. Standard errors clustered by school and cohort are in parentheses. TVA is estimated using the baseline control vector described in the text. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. We report the coefficient of TVA on the rank of the attending institution by reversing the regression sign. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A15: Main Estimates on Test Scores with Additional Controls and Alternative Interactions between the Various Fixed Effects

| | Subject-Specific Test Scores, Stacked $11^{th}$and $12^{th}$ Grades, 2003-2005 | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.205 | 0.209 | 0.204 | 0.205 | 0.208 | 0.212 |
| | (0.021)*** | (0.024)*** | (0.022)*** | (0.020)*** | (0.020)*** | (0.020)*** |
| $N$ | 42,731 | 39,282 | 42,712 | 42,733 | 42,733 | 42,733 |
| Controls as in Table 5, column 3, Panel A | ✓ | ✓ | ✓ | ✓ | | |
| Class FE × Student FE × Grade FE | | ✓ | | | | |
| Grade FE × Student FE × Year FE | | | ✓ | | | |
| Controls as in Table 5, column 3, Panel A, but replace Class FE with School FE | | | | ✓ | ✓ | ✓ |
| Core Subjects Indicator | | | | | ✓ | |
| Grade FE × School FE × Year FE | | | | | | ✓ |
| | Subject-Specific Test Scores $12^{th}$ Grade, 2003-2011 | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.192 | 0.200 | 0.192 | 0.191 | 0.191 | 0.191 |
| | (0.040)*** | (0.042)*** | (0.040)*** | (0.039)*** | (0.039)*** | (0.039)*** |
| $N$ | 38,244 | 35,846 | 38,244 | 38,244 | 38,244 | 38,244 |
| Controls as in Table 5, column 4, Panel A | ✓ | ✓ | ✓ | ✓ | | |
| Class FE × Student FE | | ✓ | | | | |
| Student FE × Year FE | | | ✓ | | | |
| Controls as in Table 5, column 4, Panel A, but replace Class FE with School FE | | | | ✓ | ✓ | ✓ |
| Core Subjects Indicator | | | | | ✓ | |
| School FE × Year FE | | | | | | ✓ |

Notes: The dependent variable is the student subject-specific test scores in $11^{th}$ and $12^{th}$ grades in 2003-2005 (Panel A) and in $12^{th}$ grade in 2003-2011 (Panel B). Table 5, column 3, Panel A includes controls for teacher characteristics, student FE, subject FE, year FE, class FE, grade FE, track FE, and a student's subject-specific previous-year test scores. Table 5, column 4, Panel A includes controls for teacher characteristics, student FE, subject FE, year FE, class FE, track FE, and a student's subject-specific previous-year test scores. Core Subjects Indicator is a binary indicator that takes the value 1 if the subject is a core subject and 0 if the subject is a track subject. Teacher characteristics include teacher gender and experience. Standard errors clustered by school and cohort are in parentheses. There are three tracks students can follow in $11^{th}$ and $12^{th}$ grades: classics, science, and exact science. All pair interactions include the main effects and all double effects between the fixed effects. Scores are standardized and have a mean of 0 and a standard deviation of 1. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A16: Relationship between the Two Long-Run Outcomes TVA

| Outcome: Teacher VA based on Degree Quality | | |
|---|---|---|
| | All Teachers | |
| Teacher VA based on University Admissions Score | 0.505<br>(0.050)*** | 0.515<br>(0.055)*** |
| | Male Teachers | |
| Teacher VA based on University Admissions Score | 0.570<br>(0.046)*** | 0.594<br>(0.051)*** |
| | Female Teachers | |
| Teacher VA based on University Admissions Score | 0.415<br>(0.075)*** | 0.403<br>(0.082)*** |
| | Experience below Median | |
| Teacher VA based on University Admissions Score | 0.417<br>(0.066)*** | 0.422<br>(0.072)*** |
| | Experience above Median | |
| Teacher VA based on University Admissions Score | 0.437<br>(0.073)*** | 0.468<br>(0.096)*** |
| Teacher FE | ✓ | ✓ |
| Year FE | | ✓ |

Notes: Each row reports coefficients from separate OLS regressions of the outcome variable (long-run TVA based on degree quality) and the other long-run TVA based on the university admissions score. Standard errors clustered by school and cohort are in parentheses. TVA is estimated using the baseline control vector described in the text. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A17: The Effect of Behavioral Teacher Value Added on University Outcomes

| | University Admissions Score 2003-2011 | Postsecondary Degree Quality 2003-2011 | Rank of Attending Institution on Degree Preference 2003-2011 | Enrollment in Postsecondary Schooling (0/1) 2003-2011 | Academic University Vs Technical School 2003-2011 | Winning a State Government Scholarship 2003-2011 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA based on Unexcused Absences | -0.130 | -0.386 | -0.242 | -0.006 | -0.011 | -0.003 |
| | (0.036)*** | (0.245) | (0.198) | (0.005) | (0.005)** | (0.002) |
| Student Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: The same regression generates all estimated coefficients in each column. Standard errors clustered by class are reported in parentheses. The treatment variable is behavioral TVA, which measures teachers' effectiveness in reducing students' unexcused absences. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so that we have one observation for each student. Previous-Year Test Scores is the $10^{th}$-grade GPA. All variables are measured in the 2003-2011 period. In column 1, the dependent variable is a student's university admissions score. In column 2, the dependent variable is a degree's quality based on annual degree admissions cutoffs. In column 3, the dependent variable is the enrolled degree's ranking on students' preference list. In column 4, the dependent variable is a binary indicator that takes the value 1 if a student enrolls in some postsecondary institution and 0 otherwise. In column 5, the dependent variable is a binary indicator for whether a student is admitted to an academic university vs a technical school. In column 6, the dependent variable is a binary indicator for whether a student receives a merit scholarship for outstanding performance from the State Government Scholarship. We report the estimated coefficient of TVA on the rank of the attending institution by reversing the regression sign. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A18: The Effect of Test-score TVA, Long-Run TVA, and Behavioral TVA on University Outcomes

| | Panel A | | | | | |
|---|---|---|---|---|---|---|
| | University Admissions Score 2003-2011 | Postsecondary Degree Quality 2003-2011 | Rank of Attending Institution on Degree Preference 2003-2011 | Enrollment in Postsecondary Schooling (0/1) 2003-2011 | Academic University Vs Technical School 2003-2011 | Winning a State Government Scholarship 2003-2011 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Average TVA based on Test Scores | 1.519 | 12.265 | 2.886 | 0.088 | 0.124 | 0.027 |
| | (0.212)*** | (1.833)*** | (0.641)*** | (0.022)*** | (0.028)*** | (0.007)*** |
| Average TVA based on Quality of Enrolled Degree | 1.474 | 13.814 | -1.604 | 0.149 | 0.051 | 0.073 |
| | (0.861)* | (6.504)** | (2.456) | (0.088)* | (0.130) | (0.044)* |
| Average TVA based on Unexcused Absences | -0.049 | -0.369 | -0.109 | 0.002 | -0.009 | -0.004 |
| | (0.024)** | (0.188)* | (0.097) | (0.003) | (0.004)** | (0.002)*** |

| | Panel B | | | | | |
|---|---|---|---|---|---|---|
| | University Admissions Score 2003-2011 | Postsecondary Degree Quality 2003-2011 | Rank of Attending Institution on Degree Preference 2003-2011 | Enrollment in Postsecondary Schooling (0/1) 2003-2011 | Academic University Vs Technical School 2003-2011 | Winning a State Government Scholarship 2003-2011 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Average TVA based on Test Scores | 1.554 | 11.869 | 2.724 | 0.099 | 0.115 | 0.025 |
| | (0.219)*** | (1.798)*** | (0.639)*** | (0.022)*** | (0.033)*** | (0.007)*** |
| Average TVA based on National Exam Performance | 0.983 | 13.810 | -0.118 | 0.046 | 0.100 | 0.072 |
| | (0.761) | (5.942)** | (2.279) | (0.081) | (0.118) | (0.037)* |
| Average TVA based on Unexcused Absences | -0.047 | -0.370 | -0.112 | 0.002 | -0.009 | -0.004 |
| | (0.023)** | (0.184)** | (0.101) | (0.003) | (0.004)** | (0.002)*** |
| Student Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimates from an OLS regression, with standard errors clustered by class in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so that we have one observation for each student. Previous-Year Test Scores is $10^{th}$-grade GPA. The treatment variable in Panel A is the average TVA based on test scores a student is exposed to, while in Panel B the treatment variable is the average TVA based on degree quality a student is exposed to. All variables are measured in the 2003-2011 period. In column 1, the dependent variable is a student's university admissions score. In column 2, the dependent variable is a degree's quality based on annual degree admissions cutoffs. In column 3, the dependent variable is the enrolled degree's ranking in students' preference list. In column 4, the dependent variable is a binary indicator that takes the value 1 if a student enrolls in some postsecondary institution and 0 otherwise. In column 5, the dependent variable is a binary indicator for whether a student is admitted to an academic university vs a technical school. In column 6, the dependent variable is a binary indicator for whether a student receives a merit scholarship for outstanding performance from the State Government Scholarship. We report the estimated coefficient of TVA on the rank of the attending institution by reversing the regression sign. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A19: Correlations between Test-Score TVA and Behavioral TVA

| Outcome: Teacher VA based on Test Scores | | |
|---|---|---|
| | All Teachers | |
| | (1) | (2) |
| Teacher VA based on Unexcused Absences | 0.000 | 0.000 |
| | (0.013) | (0.014) |
| | Male Teachers | |
| Teacher VA based on Unexcused Absences | -0.008 | -0.012 |
| | (0.016) | (0.017) |
| | Female Teachers | |
| Teacher VA based on Unexcused Absences | 0.008 | 0.012 |
| | (0.019) | (0.019) |
| | Experience above Median | |
| Teacher VA based on Unexcused Absences | 0.003 | 0.007 |
| | (0.015) | (0.011) |
| | Experience below Median | |
| Teacher VA based on Unexcused Absences | 0.008 | 0.007 |
| | (0.019) | (0.020) |
| Teacher FE | ✓ | ✓ |
| Year FE | | ✓ |

Notes: Each row reports coefficients from separate OLS regressions of the outcome variable (test-score TVA) on TVA calculated based on unexcused absences. Standard errors clustered by school and cohort are in parentheses. TVA is estimated using the baseline control vector described in the text. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A20: Heterogeneity of Teacher Value Added by Student Previous-Year Test Scores

| | (1) | (2) | (3) |
|---|---|---|---|
| Teacher VA × Student Previous-Year Test Scores | 0.016 | 0.017 | 0.040 |
| | (0.013) | (0.014) | (0.014)*** |
| Teacher VA | 0.238 | 0.229 | 0.204 |
| | (0.024)*** | (0.025)*** | (0.021)*** |
| Student Previous-Year Test Scores | 0.671 | 0.666 | 0.196 |
| | (0.010)*** | (0.010)*** | (0.010)*** |
| $N$ | 42,734 | 42,732 | 42,731 |
| Subject, Track, Grade, Year FE | ✓ | ✓ | ✓ |
| Class FE | | ✓ | ✓ |
| Student FE | | | ✓ |

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels reported in parentheses. We have one observation for each student-grade-subject-school-class-year cell. The previous-year test score refers to the same subject for which the TVA is calculated. The outcome variable is the subject-specific test score, while we stack data for the $11^{th}$ and $12^{th}$ grades. *** p<0.01, ** p<0.05, * p<0.1

Table A21: The Effect of Teacher Value Added on High School Outcomes by Teacher Gender

| | Subject-specific National Scores $11^{th}$ Grade | | Subject-specific National Scores $12^{th}$ Grade | | Subject-specific National Scores, Stacked $11^{th}$ + $12^{th}$ Grades | |
|---|---|---|---|---|---|---|
| | Males | Females | Males | Females | Males | Females |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.200 | 0.197 | 0.254 | 0.167 | 0.188 | 0.194 |
| | (0.042)*** | (0.048)*** | (0.057)*** | (0.052)*** | (0.027)*** | (0.037)*** |
| *N* | 12,591 | 10,552 | 8,849 | 9,867 | 21,778 | 20,790 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | | | | | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | | | | | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels reported in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. We have one observation for each student-grade-subject-school-class-year cell. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector described in the text. Teacher characteristics include gender (1=female) and experience. Data from the 2003-2005 sample are used in this table. *** p<0.01, ** p<0.05, * p<0.1

Table A22: The Effect of Teacher Value Added on High School Outcomes by Student Gender

| | Subject-specific National Score $11^{th}$ Grade | | Subject-specific National Score $12^{th}$ Grade | | Subject-specific National Score, Stacked $11^{th}+12^{th}$ Grades | |
|---|---|---|---|---|---|---|
| | Males | Females | Males | Females | Males | Females |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.213 | 0.193 | 0.157 | 0.257 | 0.208 | 0.203 |
| | (0.032)*** | (0.035)*** | (0.042)*** | (0.044)*** | (0.025)*** | (0.027)*** |
| $N$ | 10,229 | 13,326 | 8,530 | 10,609 | 18,777 | 23,950 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | | | | | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | | | | | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels reported in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. We have one observation for each student-grade-subject-school-class-year cell. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector described in the text. Teacher characteristics include gender (1=female) and experience. Data from the 2003-2005 sample are used in this table. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A23: The Effect of Teacher Value Added on Test Scores by Type of Subject (Core or Track)

| | Classics Core | Science Core | Exact Science Core | Classics Track | Science Track | Exact Science Track |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.143 | 0.129 | 0.122 | 0.073 | 0.191 | 0.231 |
| | (0.047)*** | (0.029)*** | (0.036)*** | (0.059) | (0.067)*** | (0.067)*** |
| *N* | 10,949 | 14,141 | 5,560 | 4,878 | 3,010 | 4,799 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | | | |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels reported in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. We use one observation for each student-grade-subject-school-class-year cell. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector described in the text. *Classics Core* includes all classics subjects from the core in both grades (modern Greek and history). *Science Core* includes all science subjects from the core in both grades (mathematics, biology, and physics in $12^{th}$ grade and algebra, geometry, and physics in $11^{th}$ grade). *Exact Science Core* includes all exact science subjects from the core in both grades (biology and physics in $12^{th}$ grade and physics in $11^{th}$ grade). Teacher characteristics include gender (1=female) and experience. Data from the 2003-2005 sample are used in this table. *** p<0.01, ** p<0.05, * p<0.1

Table A24: Heterogeneity of Teacher ValueAdded by Student Previous-Year Test Scores and Type of Subject

| | Classics Core | Science Core | Exact Science Core | Classics Track | Science Track | Exact Science Track |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher TV × Student Previous-Year Test Scores | -0.009 | 0.053 | 0.018 | -0.052 | 0.059 | 0.068 |
| | (0.020) | (0.020)*** | (0.040) | (0.026)** | (0.032)* | (0.059) |
| Teacher VA | 0.143 | 0.129 | 0.122 | 0.069 | 0.173 | 0.232 |
| | (0.046)*** | (0.029)*** | (0.036)*** | (0.058) | (0.064)*** | (0.066)*** |
| Student Previous-Year Test Scores | 0.075 | 0.023 | -0.284 | -0.093 | 0.024 | 0.004 |
| | (0.018)*** | (0.012)* | (0.026)*** | (0.020)*** | (0.025) | (0.019) |
| *N* | 10,949 | 14,141 | 5,560 | 4,878 | 3,010 | 4,799 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimated effects from an OLS regression, with standard errors clustered at school and cohort levels reported in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out TVA estimate. We have one observation for each student-grade-subject-school-class-year cell. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector described in the text. Teacher characteristics include gender and experience based on previous teaching workload in the 2003-2005 sample. *Classics Core* subjects include all classics subjects from the core in both grades (modern Greek and history). *Science Core* subjects include all science subjects from the core in both grades (mathematics, biology, and physics in $12^{th}$ grade and algebra, geometry, and physics in $11^{th}$ grade). *Exact Science Core* subjects include all exact science subjects from the core in both grades (biology and physics in $12^{th}$ grade and physics in $11^{th}$ grade). Teacher characteristics include gender (1=female) and experience. Data from the 2003-2005 sample are used in this table. *** p<0.01, ** p<0.05, * p<0.1

Table A25: The Effect of Teacher Value Added on Test Scores by Type of Subject (Core or Track) and Student Gender

| | Classics Core | | Science Core | | Exact Science Core | | Classics Track | | Science Track | | Exact Science Track | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Student Gender: | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females | Males | Females |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Teacher VA | 0.181 | 0.111 | 0.096 | 0.153 | 0.061 | 0.168 | 0.093 | 0.067 | 0.184 | 0.203 | 0.210 | 0.274 |
| | (0.058)*** | (0.051)** | (0.028)*** | (0.039)*** | (0.036)* | (0.046)*** | (0.106) | (0.061) | (0.081)** | (0.070)*** | (0.062)*** | (0.088)*** |
| *N* | 4,849 | 6,100 | 6,261 | 7,880 | 2,485 | 3,075 | 954 | 3,918 | 1,447 | 1,561 | 2,957 | 1,839 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimated effects from an OLS regression, with standard errors clustered at school and cohort levels in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so that we have one observation for each student-subject-school-year in all regressions. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector, which includes previous-year's own-subject test scores, student-level characteristics including age, gender, a binary indicator for being born in the first quarter of the birth year, class size, school-grade enrollment, and grade and year dummies. *Classics Core* includes all classics subjects from the core in both grades (modern Greek and history). *Science Core* includes all science subjects from the core in both grades (mathematics, biology, and physics in $12^{th}$ grade and algebra, geometry, and physics in $11^{th}$ grade). *Exact Science Core* includes all exact science subjects from the core in both grades (biology and physics in $12^{th}$ grade and physics in $11^{th}$ grade). Teacher characteristics include teacher gender (1=female) and experience. Data from the 2003-2005 sample are used in this table. *** p<0.01, ** p<0.05, * p<0.1

Table A26: The Effect of Teacher Value Added on Test Scores by Type of Subject in Core Subjects, Student Gender, and Teacher Gender

| Subject Type | Classics | | Classics | | Science | | Science | | Exact Science | | Exact Science | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Teacher Gender | Male | | Female | | Male | | Female | | Male | | Female | |
| Student Gender | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Teacher VA | 0.180 | 0.361 | 0.162 | 0.011 | 0.060 | 0.122 | 0.083 | 0.155 | 0.051 | 0.178 | 0.134 | 0.087 |
| | (0.172) | (0.155)** | (0.101) | (0.070) | (0.045) | (0.047)** | (0.028)*** | (0.060)** | (0.062) | (0.053)*** | (0.078) | (0.043)* |
| *N* | 1,446 | 1,754 | 2,885 | 3,645 | 3,514 | 4,577 | 2,465 | 2,964 | 1,160 | 1,423 | 1,140 | 916 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimated effects from an OLS regression, with standard errors clustered at school and cohort levels in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out TVA estimate. The dataset is stacked, so that we have one observation for each student-subject-school-year in all regressions. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector, which includes previous-year own-subject test scores, student-level characteristics including age, gender, a binary indicator for being born in the first quarter of the birth year, class size, school-grade enrollment, and grade and year dummies. In this table we use only core subjects. Subject Type of *Classics* includes all classics subjects from the core in both grades (modern Greek and history). Subject Type of *Science* includes all science subjects from the core in both grades (mathematics, biology and physics in $12^{th}$ grade and algebra, geometry, and physics in $11^{th}$ grade). Subject Type of *Exact Science* includes all exact science subjects from the core in both grades (biology and physics in $12^{th}$ grade and physics in $11^{th}$ grade). Teacher characteristics include teacher experience. *** p<0.01, ** p<0.05, * p<0.1

Table A27: The Effect of Teacher Value Added on Test Scores by Type of Subject in Tracks, Student Gender, and Teacher Gender

| Subject Type | Classics | | Classics | | Science | | Science | | Exact Science | | Exact Science | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Teacher Gender | Male | | Female | | Male | | Female | | Male | | Female | |
| Student Gender | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Teacher VA | 0.245 | -0.008 | 0.232 | 0.161 | 0.214 | 0.167 | 0.162 | 0.163 | 0.252 | -0.054 | 0.204 | -0.000 |
| | (0.273) | (0.108) | (0.100)** | (0.072)** | (0.176) | (0.156) | (0.128) | (0.115) | (0.103)** | (0.094) | (0.231) | (0.146) |
| *N* | 367 | 1,485 | 529 | 2,240 | 741 | 917 | 463 | 420 | 1,557 | 970 | 574 | 870 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each column reports estimated effects from an OLS regression, with standard errors clustered at school and cohort levels in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out TVA estimate. The dataset is stacked so that we have one observation for each student-subject-school-year in all regressions. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector, which includes previous-year own-subject test scores, student-level characteristics including age, gender, a binary indicator for being born in the first quarter of the birth year, class size, school-grade enrollment, and grade and year dummies. In this table we use only track subjects. Subject Type of *Classics* track subjects include ancient Greek, philosophy and Latin in $11^{th}$ grade and ancient Greek, Latin, literature, and history in $12^{th}$ grade. Subject Type of *Science* track subjects include mathematics, physics, and chemistry in $11^{th}$ grade and ancient biology, mathematics, physics, and chemistry in $12^{th}$ grade. Subject Type of *Exact Science* track subjects include mathematics, physics, and computer science in $11^{th}$ grade and ancient biology, mathematics, physics, business administration, and computer science in $12^{th}$ grade. Teacher characteristics include teacher experience. *** p<0.01, ** p<0.05, * p<0.1

Table A28: The Effect of Teacher Value Added on Test Scores by Teacher Experience and Teacher Gender

| | Experience>=12 | Experience<12 | | | | |
|---|---|---|---|---|---|---|
| Teacher Gender: | All | All | Females | | Males | |
| | | | Exper.>=10 | Exper.<10 | Exper.>=14 | Exper.<14 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.211 | 0.196 | 0.127 | 0.261 | 0.231 | 0.111 |
| | (0.041)*** | (0.028)*** | (0.055)** | (0.071)*** | (0.035)*** | (0.035)*** |
| *N* | 21,628 | 20,977 | 10,782 | 10,674 | 10,372 | 9,501 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: The median experience of a teacher is 12 classes. The median experience for female teachers is 10 classes and for male teachers 14 classes. Each column reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so that we have one observation for each student-subject-school-year in all regressions. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector, including previous-year's own-subject test scores, student-level characteristics including age, gender, class size, school-grade enrollment, and grade, year, and subject dummies. In all regressions, we additionally control for teacher gender (1=female) and previous-year test scores in the same subject. Data from the 2003-2005 sample are used in this table. *** p<0.01, ** p<0.05, * p<0.1

Table A29: The Effect of Teacher Value Added on Test Scores by Class Size and Teacher Gender

| | Class Size>=21 | Class Size<21 | Class Size>=21 | Class Size<21 | Class Size>=21 | Class Size<21 |
|---|---|---|---|---|---|---|
| Teacher Gender: | All | All | Females | Females | Males | Males |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Teacher VA | 0.211 | 0.196 | 0.127 | 0.261 | 0.231 | 0.111 |
| | (0.041)*** | (0.028)*** | (0.055)** | (0.071)*** | (0.035)*** | (0.035)*** |
| *N* | 21,628 | 20,977 | 10,782 | 10,674 | 10,372 | 9,501 |
| Teacher Characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: The mean class size is 21 students. Each column reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out teacher VA estimate. The dataset is stacked so we have one observation for each student-subject-school-year in all regressions. TVA is scaled in student test-score standard deviations and estimated using data from classes taught by the same teacher in all other classes and years in the sample. TVA is estimated using the baseline control vector, which includes previous-year's own-subject test scores, student-level characteristics including age, gender, an indicator variable for being born in the first quarter of the birth year, and class size, school-grade enrollment, and school-grade and year dummies. In all regressions, we additionally control for teacher gender (1=female) teacher experience, and previous-year test scores in the same subject. Data from the 2003-2005 sample are used in this table. *** p<0.01, ** p<0.05, * p<0.1

Table A30: The Effect of Teacher VA in Closest Subject on Students' Probability to Study in a University Department that is the Natural Follow-up of School Track for Full Sample and by Gender, 2003-2011

| | Indicator if Choice of Department is a Natural Follow-up of School Track | | | |
|---|---|---|---|---|
| | All Tracks | | | Science and Exact Science Track |
| | (1) | (2) | (3) | (4) |
| Panel A: Full Sample | 0.043 | 0.045 | 0.045 | 0.073 |
| | (0.020)** | (0.021)** | (0.021)** | (0.026)*** |
| $N$ | 1,495 | 1,495 | 1,495 | 957 |
| Panel B: Females | 0.035 | 0.036 | 0.036 | 0.066 |
| | (0.027) | (0.027) | (0.027) | (0.039)* |
| $N$ | 876 | 876 | 876 | 451 |
| Panel C: Males | 0.053 | 0.055 | 0.055 | 0.072 |
| | (0.032)* | (0.033)* | (0.033)* | (0.036)** |
| $N$ | 619 | 619 | 619 | 505 |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| School FE | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | |
| Student Characteristics | | ✓ | ✓ | ✓ |
| Average Teacher Characteristics | | ✓ | ✓ | ✓ |
| Average Class Size | | | ✓ | ✓ |

Notes: Each cell presents OLS estimates from a different regression. The outcome variable in columns 1-3 is a binary indicator that takes the value 1 if students enroll in a university field (exact science, science, humanities, social science) equivalent to their high school track (exact science, science, humanities) and 0 otherwise. The dependent variable in column 4 is a binary indicator that takes the value 1 if students enroll in an university field (only exact science or science) equivalent to their high school track (exact science, science) and 0 otherwise. TVA is calculated as the average TVA in the closest high school subjects to the student's university department study. The average track TVA is used whenever there is no exact subject correspondence. We use the average track TVA in the exact science high school track for social science university degrees, since most students who enroll in social science degrees follow the exact science track in high school. The subjects we use for each field of study are the following: for economics we use the TVA in economics in $12^{th}$ grade. For business, we use the TVA in business administration in the track in $12^{th}$ grade. For history, we use the TVA in history in the track in $12^{th}$ grade. For mathematics, we use the average TVA in mathematics in the track in $12^{th}$ grade. For physics, we use the average TVA in physics in the track in $12^{th}$ grade. For engineering, we use the average TVA in physics and biology in the track in $12^{th}$ grade. For computer science, we use the TVA in computer science in the track in $12^{th}$ grade. For health-related fields (medicine, dentistry, veterinary, and pharmacy), we use the average TVA in the science or exact science track in $12^{th}$ grade. We use the average TVA in the classics track in $12^{th}$ grade for the remaining humanities departments. We use the average TVA in the exact science track in $12^{th}$ grade for the remaining exact science departments. We use the TVA in biology in the science track in $12^{th}$ grade for the remaining science departments. Student characteristics include controls for student age and gender (=1 if female). Teacher characteristics include controls for teacher gender (=1 if female) and experience. Previous-Year Test Scores control for $10^{th}$ grade GPA. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A31: The Effect of Teacher Value Added on Test Scores by Teacher Value Added Quintiles

| | **All Subjects** | **Classics Core** | **Science Core** | **Classics Track** | **Science Track** | **All Classics** | **All Science** |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Second Quintile from Bottom of TVA | 0.105*** | 0.123** | 0.065** | 0.010 | 0.047 | 0.096** | 0.072*** |
| | (0.027) | (0.053) | (0.026) | (0.053) | (0.052) | (0.041) | (0.024) |
| Middle Quintile of TVA | 0.131*** | 0.093** | 0.093*** | 0.015 | 0.048 | 0.079** | 0.102*** |
| | (0.029) | (0.038) | (0.030) | (0.061) | (0.078) | (0.033) | (0.029) |
| Second Quintile from Top of TVA | 0.202*** | 0.194*** | 0.148*** | 0.044 | 0.138* | 0.148*** | 0.147*** |
| | (0.027) | (0.046) | (0.037) | (0.071) | (0.070) | (0.040) | (0.029) |
| Top Quintile of TVA | 0.260*** | 0.143** | 0.230*** | 0.173** | 0.220** | 0.180*** | 0.229*** |
| | (0.029) | (0.063) | (0.040) | (0.078) | (0.086) | (0.056) | (0.032) |
| Observations | 36,347 | 10,949 | 15,613 | 4,879 | 4,408 | 15,884 | 20,362 |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: The same regression generates all estimates coefficients in each column. Teachers are ranked based on their TVA independent of which subject they teach. We show the estimated effects for each TVA quintile, except for the first (bottom-quality) quintile, which is omitted as a point of comparison. The table presents the estimated effects of different quintiles of TVA on student test scores in different subjects (shown in the column headings) for the 2003-2005 period. Data for both grades are used and grade fixed effects are added. Student test scores are standardized at school, year, and grade level. *All Subjects* includes all core and track subjects in both grades. *Classics Core* and *Classics Track* include classics-related subjects from the core (modern Greek and history) and all subjects in the classics track in both $11^{th}$ and $12^{th}$ grades. *Science Core* and *Science Track* include the science-related subjects from the core in both grades (mathematics, biology, and physics in $12^{th}$ grade and algebra, geometry, and physics in $11^{th}$ grade) and all subjects in the science track in both grades. *Exact Science Core* and *Exact Science Track* include the exact science-related subjects from the core in both grades (biology and physics in $12^{th}$ grade and physics in $11^{th}$ grade) and all subjects in the exact science track in both grades. Standard errors are clustered at the school and cohort levels. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table A32: Correlation Between Absenteeism, Student Characteristics and Previous-Year Test Scores

| | Outcome Variable: Excused Absences | Outcome Variable: Unexcused Absences |
|---|---|---|
| | (1) | (2) |
| Age | 1.007 | 0.076 |
| | (0.853) | (0.333) |
| Female Student | 3.141 | -0.742 |
| | (0.807)*** | (0.451) |
| Previous-Year Test Scores | 0.526 | -1.071 |
| | (0.229)** | (0.117)*** |
| $R^2$ | 0.34 | 0.55 |
| Grade FE | ✓ | ✓ |
| Year FE | ✓ | ✓ |
| School FE | ✓ | ✓ |
| Observations | 2,870 | 2,870 |

Notes: Each column reports coefficients from an OLS regression with standard errors clustered at school and cohort levels reported in parentheses. Regressions are run on the sample used to estimate the baseline TVA model, restricted to observations with a non-missing leave-out TVA estimate. *** p<0.01, ** p<0.05, * p<0.1

Table A33: The Effect of Absences on Test Scores by Type of Subjects—Classics, Science, and Exact Science

| | All | Classics Core | Classics Track | Science Core | Science Track | Exact Science Core | Exact Science Track |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Panel A: Total Absences | -0.000 | 0.000 | -0.001 | -0.000 | -0.001 | -0.001 | -0.003 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001)* |
| *N* | 42,731 | 10,949 | 4,878 | 14,141 | 3,010 | 5,560 | 4,799 |
| Panel B: Excused Absences | 0.000 | 0.001 | 0.000 | -0.000 | -0.001 | -0.001 | -0.002 |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001)* |
| *N* | 42,731 | 10,949 | 4,878 | 14,141 | 3,010 | 5,560 | 4,799 |
| Panel C: Unexcused Absences | -0.002 | -0.002 | -0.004 | -0.002 | 0.000 | -0.002 | -0.002 |
| | (0.001) | (0.002) | (0.002)* | (0.001) | (0.002) | (0.001) | (0.002) |
| *N* | 42,731 | 10,949 | 4,878 | 14,141 | 3,010 | 5,560 | 4,799 |
| Teacher Characteristics including TVA | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Previous-Year Test Scores | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Grade FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Track FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Class FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Each row reports coefficients from an OLS regression, with standard errors clustered at school and cohort levels reported in parentheses. *Classics Core* includes all classics-related subjects from the core (modern Greek and history) in both grades. *Classics Track* includes all subjects in the classics track in both grades. *Science Core* includes all science-related subjects from the core in both grades (mathematics, biology, and physics in $12^{th}$ grade and algebra, geometry, and physics, in $11^{th}$ grade). *Science Track* includes all subjects in the science track in both grades. *Exact Science Core* includes all exact science-related subjects from the core in both grades (biology and physics in $12^{th}$ grade and physics in $11^{th}$ grade). *Exact Science Track* includes all subjects in the exact science track in both grades. *Teacher Characteristics* include teacher gender, experience and teacher VA. Data for the period 2003-2005 are used. *** p<0.01, ** p<0.05, * p<0.1

Table A34: Descriptive Statistics for the Sample Used in the Retention Analysis

| | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| ***Panel A*** | | | | | |
| New Teacher Indicator | 1,267 | 0.271 | 0.445 | 0 | 1 |
| Exit Teacher Indicator | 1,267 | 0.268 | 0.443 | 0 | 1 |
| Avg Number of Years a Teacher is Observed in the Period 2003-2011 | 1,267 | 4.763 | 2.083 | 2 | 9 |
| School Ranking | 1,267 | 10.060 | 5.768 | 1 | 21 |
| Female Teacher | 1,267 | 0.514 | 0.500 | 0 | 1 |
| Female School Principal | 1,267 | 0.000 | 0.000 | 0 | 0 |
| ***Panel B*** | | | | | |
| Mean TVA for All Teachers | 927 | -0.167 | 0.718 | -7.333 | 2.300 |
| Mean TVA for New Teachers | 343 | -0.104 | 0.630 | -2.825 | 1.998 |
| Mean TVA of Other Teachers in the School | 584 | -0.203 | 0.763 | -7.333 | 2.300 |
| TVA of Last Year for Exiting Teachers | 276 | -0.116 | 0.693 | -2.577 | 2.300 |
| ***Panel C*** | | | | | |
| ***High Achieving Schools*** | | | | | |
| Entry Rate | 587 | 0.267 | 0.443 | 0 | 1 |
| Exit Rate | 587 | 0.276 | 0.447 | 0 | 1 |
| TVA | 425 | -0.102 | 0.790 | -7.333 | 2.300 |
| ***Low Achieving Schools*** | | | | | |
| Entry Rate | 680 | 0.274 | 0.446 | 0 | 1 |
| Exit Rate | 680 | 0.262 | 0.440 | 0 | 1 |
| TVA | 502 | -0.221 | 0.647 | -2.825 | 1.998 |
| ***Panel D*** | | | | | |
| Retention in t+1 | 927 | 0.701 | 0.458 | 0 | 1 |

Notes: TVA is teacher-, and year-specific. *New Teacher Indicator* is a binary indicator that takes the value 1 if a teacher is new to the school and 0 otherwise. *Exit Teacher Indicator* is a binary indicator that takes the value 1 if a teacher exits the school and 0 otherwise. TVA cannot be calculated in the year the teacher leaves the school. School ranking ranges from 1 to 21 and measures school quality, in which a school takes the value 0 for the lowest-performing high school and 21 for the highest-performing high school. *Female Teacher* is a binary indicator that takes the value 1 if a teacher is female and 0 otherwise. *Female School Principal* is a binary indicator that takes the value 1 if the high school principal is female and 0 otherwise. *High-Achieving Schools* have rankings above the median (10). *Low-Achieving Schools* have rankings below or equal to the median (10). Panel D shows descriptive statistics for the retention status used in Section 7. Data from 2004 to 2011 are used to calculate TVA. School ranking is calculated using the 2003 data. Data from 21 schools are used.