I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13146

Determinants of Teacher Value-Added in Public Primary Schools: Evidence from Administrative Panel Data

Ryuichi Tanaka
Shun-ichiro Bessho
Akira Kawamura
Haruko Noguchi
Koichi Ushijima

APRIL 2020

DISCUSSION PAPER SERIES

IZA DP No. 13146

Determinants of Teacher Value-Added in Public Primary Schools: Evidence from Administrative Panel Data

Ryuichi Tanaka
University of Tokyo and IZA

Shun-ichiro Bessho
University of Tokyo

Akira Kawamura
Kanagawa University of Human Services

Haruko Noguchi
Waseda University

Koichi Ushijima
University of Tsukuba

APRIL 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Determinants of Teacher Value-Added in Public Primary Schools: Evidence from Administrative Panel Data*

This study estimates teacher value-added (TVA) for language arts and mathematics test scores of students in public primary schools to investigate the empirical relationship between testscore TVA and observable traits and promotions of teachers. Our empirical strategy employs Chetty, Friedman, and Rockoff (2014a) with school-year fixed effects as an additional control for potential sorting of students across schools. Using unique administrative panel data of students in public primary schools of a large municipality of Japan, we find TVA distribution to have variance comparable to ones observed in the U.S. schools. Using TVA estimates, we examine their associations with gender, teaching experience, age, and promotions of teachers. We find that these observable characteristics of teachers are statistically significantly associated with TVA estimates. Additionally, we find that TVA estimates are positively associated with teacher promotions.

JEL Classification: H75, I21, J24, J45

Keywords: education, teacher value-added, class size, teaching experience, promotion

Corresponding author:
Ryuichi Tanaka
Institute of Social Science
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033
Japan
E-mail: ryuichi.tanaka@iss.u-tokyo.ac.jp

* We are grateful to valuable comments from Hideo Akabayashi, Hidehiko Ichimura, Takao Kato, Daiji Kawaguchi, Nobuyoshi Kikuchi, Toru Kitagawa, Victor Lavy, Ryo Nakajima, Masao Ogaki, Fumio Ohtake, Jeffrey Smith, Katsuya Takii, and participants of TPLS 2019 and AASLE 2019. We received generous support from a local municipality providing the data for this study. This research is supported by the Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (grant numbers 16H03636, 19H00602, and 20H00069). Of course, all remaining errors are solely ours.

1. Introduction

Teachers are among the most important educational resources at any school level. It is widely accepted that teachers impact students' academic achievements: they tutor students directly, teach academic contents such as language arts and mathematics, and manage schools and classroom environments that are extremely important for student learning. Hence, it is very important to keep highly qualified teachers and to recruit high quality teachers in order to improve students' educational achievement.

To find effective teachers, it is imperative to measure individual teaching quality. Typical observable measurements of teachers' effectiveness include teaching experience; teaching certificate; teachers' educational backgrounds, such as college major and degree; and teachers' test scores. However, there are ample unobservable characteristics of teachers that impact students such as pedagogy, classroom management skill, teaching philosophy, communication skill, motivation, individual traits and preferences. As a matter of fact, many researchers report that these observable characteristics of teachers have limited power to predict teachers' subsequent classroom performance to improve students' academic ability, as measured by test scores (e.g., Hanushek 2001 and Hanushek and Rivkin 2006).

An alternative approach to measure teacher quality is to estimate "total teacher effects" by estimating teacher value-added (TVA) on specific outcome variables such as language arts and mathematics test scores. TVA is defined as contribution of each teacher to growth of test scores of individual student in a given year. It is typically measured as average contribution to students taught by each teacher and used to compare to the teaching performance of other teachers. A variety of studies have pursued this approach, such as Hanushek (1971, 1992), Rockoff (2004), Rivkin, Hanushek and Kain (2005), Hanushek et al. (2005), Kane and Staiger (2008), and Chetty, Friedman, and Rockoff (2014a) with students' panel data in the United States.[1]

Although estimated TVA is one of predictors of both short- and long-term student outcomes as shown in Chetty, Friedman, and Rockoff (2014b) in the U.S. context, outcome information is often unavailable due to lack of productive observation or measurement in majority of non-U.S. countries and even in the U.S. states. Most of studies cited above are for the schools in the U.S., and TVA estimates are relatively scarce in non-U.S. countries. The lack of information on teacher effectiveness could be a serious concern, particularly when schools

[1] For other examples of TVA estimation in the United States, see Nye, Konstantopoulos, and Hedges (2004) for Tennessee; Jacob and Lefgren (2008) for an undisclosed city; Kane, Rockoff and Staiger (2008) for New York City; Koedel and Betts (2011) for San Diego; Rothstein (2010) and Jackson (2018) for North Carolina; Anderson, Barrow, and Sander (2007) for Chicago; Harris and Sass (2011) for Florida; and Bacher-Hick, Kane, and Staiger (2014) for Los Angeles.

and board of education need to appoint new teachers. In the absence of accurate information about new teacher quality, schools and educational administrators must take risks to expose students to yet-unqualified new teachers for several years. In order to avoid taking these risks, they might seek any observable trait of teachers useful to predict their teaching quality.[2]

The primary purpose of this study is to investigate the empirical relationship between test-score TVA and observable traits and promotions of teachers to evaluate whether observable characteristics of teachers deliver useful information when educational administrators recruit and promote effective teachers. To achieve this goal, we focus on a large municipality in Tokyo metropolis of Japan where all schools are managed under a common local education policy set by the municipal board of education. We obtained seven-year's administrative data from the municipality and constructed unique panel data of all students from second to fifth grades in public primary schools of the municipality. The use of this unique dataset enables us to estimate TVA in schools under homogenous environment of local education policy. More importantly, thanks to teacher rotation system in Japan explained in detail in the next section, endogeneity of matching between teachers and students is less severe in Japanese schools than schools in other countries such as the United States.

Estimating TVA applying the method proposed by Chetty, Friedman, and Rockoff (2014a) with school-year fixed effects as an additional control for the potential sorting of students across schools, as addressed by Chetty, Friedman, and Rockoff (2016), we find the following. First and most importantly, variation of TVA among teachers is substantial: improvement of TVA by one standard deviation raises student's test score by 0.150-0.155 standard deviation for language arts and by 0.174-0.179 standard deviation for mathematics. These estimates are comparable to those in Chetty, Friedman, and Rockoff (2004a) for public primary school students in districts in the United States. Second, we find that TVA is highly persistent over years: our estimates reveal that the first-order autocorrelation of TVA for language arts test scores is 0.490, and that for mathematics is 0.575.

To validate our TVA estimates, we conduct additional analyses proposed by Chetty, Friedman, and Rockoff (2004a). First, we confirm that the estimated TVA is the best linear predictor of test score residuals by regressing residual test score on the estimated TVA. Second, we confirm that endogenous matching between students and teachers is not a serious concern for the estimation of TVA in our data. We regress test score residuals predicted from their past

[2] Some of school districts in the U.S. (e.g., Chicago public schools, New York City's department of education and District of Columbia public schools) had adopted teacher value-added to decide on issues of teacher retention and the awarding of bonuses, but the number of such applications of TVA is limited in practice.

growth on the estimated TVA and find no statistically significant relationship, which would have correlated if matching between students and teachers were based on past growth of test scores. The results of these analyses support the validity of our estimates of TVA.

Using the estimated TVA, we examine how it is associated to observable teacher characteristics such as gender, teaching experience, and age. Since both language arts and mathematics are taught by the same classroom teacher in public primary schools in Japan, we estimate the association between the estimated TVA and gender, teaching experience, and teacher age for each subject considering seemingly unrelated correlations of error terms between subjects. We find female teachers tended to have higher value-added for language arts than male teachers, but there is no systematic gender difference in math TVA. Teaching experience is positively associated to both language arts and mathematics TVA, conditional on teacher's age. Similarly, teachers' ages are negatively associated with TVA, conditional on years of teaching experience.

Furthermore, we explore the relationship between these observable characteristics and the TVA estimates by decile of TVA distribution. We conduct quantile regression to reveal the heterogenous relation between the estimated TVA and observable teacher characteristics by decile of TVA. We find that the relationship between these observables and TVA is roughly uniform across the distribution of TVA for language arts. However, we also find that for mathematics, teaching experience is positively associated with the estimated TVA only for teachers with relatively low value-added.

The paper's findings regarding the relationship between the TVA and observable characteristics of teachers indicate that gender, teaching experience, and teacher age can be used to detect high-performing teachers. However, although some observable characteristics account for the variation of TVA, we find high cross-subject correlation of TVA even after removing the effects of these observable characteristics common to both subjects. The raw correlation coefficient of the estimated value-added between language arts and mathematics is 0.6603, and the correlation coefficient of the residual from seemingly unrelated regression remains 0.6556: only 0.7% of the cross-subject correlation of TVA is explained by gender, teaching experience, and age of teachers. These results indicate that it remains difficult to find a good teacher based solely on these limited numbers of observable teacher characteristics.

Finally, we examine the relationship between the estimated TVA and the promotion probability of teachers. In the institutional setting from which our data derive, TVA has been neither calculated nor used for teacher evaluation. However, it may be possible for members of boards of education and school principals to have a sense of TVA (see Jacob and Lefgren,

2008) and to use this information to evaluate teachers' performance implicitly. In terms of teacher promotion, we find that the TVA for language arts is statistically significantly associated with the promotion probability to chief teacher (second rank from the teacher base), but no relationship is found for the probability of promotion to managing teacher (third rank from the teacher base). These results support the hypothesis that the board of education members and school principals have information related to TVA and that they use it for evaluation of teachers for promotion to chief teacher even without formal TVA measurement.

Our paper contributes to a strand of existing literature on the estimation of teacher's impact on student achievement by adding evidence showing the robustness of TVA estimates in a different context. For example, Hanushek and Rivkin (2010) reports the summary of several estimates of TVA in the United States. One standard deviation of estimated teacher effectiveness, measured by the standardized test score with zero mean and one standard deviation ranged from 0.08 to 0.26 for language arts (reading) test scores and from 0.11 to 0.36 for mathematics.[3] Our estimates of teacher effectiveness based on TVA, using new and unique panel data in Japan as another advanced country with high educational achievement shown in international evaluation such as PISA and TIMSS, are in range of teacher effectiveness estimates in the U.S., showing the robustness of our TVA estimates.

Another contribution of our paper is the demonstration of the observable teacher characteristics that are useful to determine teacher effectiveness, allowing for heterogeneity. In an examination of the relationship between TVA and teacher's observables, Bau and Das (2020) estimated TVA and estimated correlation with observable characteristics of teachers using data from Pakistan. They found that the first two years of tenure and content knowledge correlated with TVA. In the current study, we find similar correlations between teaching experience and TVA, but we additionally find that these correlations are heterogenous according to subject and the quantile of TVA distribution. In particular, we find gender differences in language arts TVA in the institutional setting of Japan, where coeducation is typical in public primary schools, to be different from schools in Pakistan, where coeducational schools are rare. Regarding the correlation between labor market compensation to teachers and TVA, Bau and Das found that wages for teachers in the public sector do not correlate with TVA. In our analysis, we find a weak but statistically significant correlation between TVA and

[3] For non-U.S. countries, estimates of teacher effectiveness are relatively scarce. Among exceptions, Azam and Kingdon (2015) estimate TVA on high school students in India. Bau and Das (2020) estimate TVA for primary school students in Pakistan. Buhl-Wiggers, Kerwin, Smith, and Thornton (2019) estimate TVA for middle school students in Uganda.

rank, that is positively correlated with teacher salary even though all teachers in our analysis are public sector teachers.

This paper contributes to the literature of education economics in its use of Japanese data. There have been several papers studying the effect of teacher characteristics and test scores using Japanese data. Hojo (2012) and Hojo and Oshio (2013) found a positive association of female teachers with math test scores of 8^{th} grade students with TIMSS data from 2007. Tanaka and Ishizaki (2018) examined the effect of teaching practice of classroom teachers on test score of students in public primary schools. They found that teaching methods focusing on logical thinking had a positive impact on students' test score of both language arts and mathematics. However, these papers do not control for teacher fixed effects or estimate TVA. As far as we know, our paper is the first to estimate TVA using student-level panel data from public primary schools in Japan.

The remainder of this paper is organized as follows. Section 2 briefly discusses the education system in Japan. Section 3 explains our empirical strategy of this study. Section 4 presents the data and descriptive statistics. Section 5 reports the estimates of teacher effectiveness and the validity of the estimated TVA. In section 6, we investigate the relationship between the estimated TVA and observable characteristics of teachers including teacher's rank. Section 7 concludes the discussion.

2. Public Primary School System in Japan

The current compulsory education system in Japan is based on the Fundamental Law of Education Act passed in 1947. Compulsory education consists of six years of primary education and three years of secondary schooling. Children who are age 6 on April 1 begin the first grade in primary school on that day. Students are assigned to a public school in their residential district. Public schooling is provided free of charge during the compulsory education period. Although students are allowed to attend a private school by paying tuition and various related costs, the majority of students attend their assigned public schools.[4]

In primary schools, each class has one classroom teacher who teaches all subjects to students in his/her class. Classroom teachers are in charge of both classroom management and subject teaching. The composition of students in a class and the assignment of a classroom teacher are subject to change when students advance to the next grade in April, but schools are allowed to maintain the same teacher/student across grades at the discretion of school principals.

[4] In 2019, 98.2% of students were enrolled in public primary schools (School Basic Survey by Ministry of Education, Culture, Sports, Science and Technology (MEXT), 2019).

The contents of teaching in Japanese schools are centralized at the national level, but their implementation is decentralized at the local municipality level. The National Central Education Council revises the School Curriculum Guideline for all subjects every ten years. However, since this guideline sets the minimum content to be taught in schools at each educational level, the local board of education has discretion in how its implementation and any additional content will be taught in schools.[5] In our analysis, focusing on one local municipality, we can avoid heterogeneity of education policy at the municipality level.

Although municipal board of education is in charge of implementation of schooling, all public primary school teachers are employed by prefectural board of education to avoid any disparity in the allocation of school resources, including teachers within a prefecture. The prefectural government decides on the allocation of teachers among municipalities based on the teacher allocation rules set by the Ministry of Education in the central government.[6] In principle, the total numbers of students and classes almost automatically determines the number of teachers allocated to each school. Each municipality could determine the allocation of teachers among schools, but there is limited room for discretion in schools' teacher allocation.

One unique feature of the allocation of teachers among schools is the teacher rotation system.[7] Japanese public schools rotate teachers and principals to different schools on a regular basis to guarantee equity of educational opportunities across school districts. As mentioned above, public school teachers are employed not by each school but by each prefecture. For this reason, the Board of Education for each prefecture holds the authority over personnel issues. While the timing of teacher transfer to another school within the same prefecture depends on the discretion of each prefecture, they are customarily transferred every 3–6 years in all prefectures.[8] Under this policy, it is difficult for students and their parents to predict teacher composition in any specific local public school over long time periods. In the municipality studied in this paper, students are allowed to choose any school within municipality when they start the first grade. However, since students are not allowed to switch schools once they start

[5] Part of the current English version of the guideline can be obtained from the website of MEXT. http://www.mext.go.jp/en/about/pablication/__icsFiles/afieldfile/2017/02/15/1374478_001.pdf (Access Date: Jan 30, 2018).

[6] Japan has three levels of administrative districts: national, prefectural, and municipal. The nation is divided into 47 prefectures. Each prefecture consists of numerous municipalities, with 1,718 in total (January 2018).

[7] The procedure of teacher reallocation is determined by "The Act on the Organization and Operation of Local Educational Administration." The procedure of implementation of teacher transfer is as follows. Firstly, school principals request it to the municipal board of education. Secondly, municipal board of education summarizes requests from school principals and reports it to prefectural board of education. Thirdly, prefectural board of education determines and approves teacher transfers after adjustments of the number of teachers allocated to municipalities.

[8] For more information on Teacher Rotation System in Japan, please visit the following website of National Institute for Educational Policy Research, Ministry of Education, Culture, Sports, Science and Technology. http://www.nier.go.jp/English/educationjapan/pdf/201703TTASJ.pdf. (Access Date: Jan 30, 2018)

the first grade without special reasons, such as bullying or parents' residential move, it is difficult for students and parents to predict who will be the classroom teacher their primary education.

Finally, students are not allowed to choose their classmates or teachers. From the student's point of view, the assignment of classroom teachers and classmates are random and to some extent conditional on their school choice. Endogenous choice of students by teachers may be a concern and even a potential threat to identification of TVA. We will discuss this potential threat in Section 5 and argue that this type of sorting is not a serious concern in our data and estimation of TVA.

3. Econometric Model

To estimate TVA, we follow the method proposed by Chetty, Friedman and Rockoff (2014a). We describe the method briefly in this section.

3.1 Regression Model

Firstly, we estimate the following linear regression model

$$\boldsymbol{A}_{it}^{*} = \boldsymbol{\beta X}_{it} + \boldsymbol{\nu}_{it} \quad (1)$$

where $\boldsymbol{A}_{it}^{*}$ is student i's educational outcome (test score) in year t. $\boldsymbol{X}_{it}$ is student i's time-varying characteristics including student's lagged test scores. $\boldsymbol{\nu}_{it}$ is unobservable "error" term. This term is decomposed into three components:

$$\boldsymbol{\nu}_{it} = \boldsymbol{\mu}_{jt} + \boldsymbol{\theta}_{c} + \tilde{\boldsymbol{\varepsilon}}_{it}$$

where $\boldsymbol{\mu}_{jt}$ is teacher j's value-added in year t, $\boldsymbol{\theta}_{c}$ is exogenous shock common to students in class c, and $\tilde{\boldsymbol{\varepsilon}}_{it}$ is idiosyncratic student-level shock. Following Chetty, Friedman and Rockoff (2014a), we assume that TVA $\boldsymbol{\mu}_{jt}$ and the composite of the other two shocks $\boldsymbol{\varepsilon}_{it} = \boldsymbol{\theta}_{c} + \tilde{\boldsymbol{\varepsilon}}_{it}$ are stationary.

The assumption of stationarity requires that mean teacher quality does not vary across years and the correlation of TVA, class-level exogenous shocks, and student-level exogenous shocks across any pair of years depends only on the amount of time that elapses between those years. This assumption reduces the number of parameters to be estimated in the estimation of TVA distribution. In particular, the variance of TVA, $\sigma_{\mu}^{2} = Var(\boldsymbol{\mu}_{jt})$, is constant across periods under the assumption of stationarity.

3.2 Estimation of Teacher Value-Added

The procedure of TVA estimation follows several steps. In the first step, we estimate the regression equation (1) by regressing test scores $\boldsymbol{A_{it}^*}$ on $\boldsymbol{X_{it}}$ and teacher fixed effect (call it $\boldsymbol{\alpha_j}$). In the second step, we compute test score residual (call it $\boldsymbol{A_{it}}$):

$$\boldsymbol{A_{it} = A_{it}^* - \widehat{\beta} X_{it}}$$

where $\boldsymbol{\widehat{\beta}}$ is the estimates of the coefficient vector in the regression equation (1) with teacher fixed effect. Note that this test score residual contains teacher fixed effect and thus equal to the estimate of $\boldsymbol{\mu_{jt} + \varepsilon_{it}}$. In the third step, we calculate average test score residual of students taught by teacher j in year t:

$$\boldsymbol{\overline{A}_{jt} = \frac{1}{n_{jt}} \sum_{i \in \{i: j(i,t) = j\}} A_{it}}$$

where $\boldsymbol{n_{jt}}$ is the number of students taught by teacher j in year t. In the fourth step, we estimate a two-sided auto-regression model of the average test score residual by OLS:

$$\boldsymbol{\overline{A}_{jt} = \sum_{s=1}^{t-1} \psi_s \overline{A}_{j,t-s} + \sum_{s=1}^{T-t} \psi^s \overline{A}_{j,t+s} + \epsilon_{jt}}$$

where T is the total years of observation and $\boldsymbol{\epsilon_{jt}}$ is white noise. In the final step, we calculate the best linear predictor of mean test score residual of students taught by teacher j in year t based on mean residual test score in prior and posterior years using the autoregression model estimated in the previous step:

$$\boldsymbol{\widehat{\mu}_{jt} = \sum_{s=1}^{t-1} \widehat{\psi}_s \overline{A}_{j,t-s} + \sum_{s=1}^{T-t} \widehat{\psi}^s \overline{A}_{j,t+s}}$$

where $\boldsymbol{\widehat{\psi}_s}$ and $\boldsymbol{\widehat{\psi}^s}$ are the OLS estimates of the coefficients of the autoregression model in the previous step. This is our measure of teacher j's value-added in year t. As emphasized in Chetty, Friedman and Rockoff (2014a), the jackknife (i.e., leaving out year t) measures of TVA are important to avoid introducing the same estimation errors on the left-hand side (that is, test score) and the right-hand side (i.e., TVA) as a cause of biases in the estimation of teacher's causal impact.

4. Data

Our primary data source is a database of students in a large urban municipality in Japan with a population of approximately 600,000. This database covers panels of students in public primary

(1st to 6th grade) and middle (7th to 9th grade) schools.[9] In this database, panel information of test scores of Japanese language arts and mathematics are available. These tests measure academic achievement in the previous grade of the understanding of basic material of Japanese Language arts and mathematics. We match this test score information with classroom teachers of the last year in the previous grade.

In addition to the test scores, information on household characteristics is available. We use student status in terms of school financial assistance receipt as a covariate to for control socioeconomic status of household of students. School financial assistance is a means-tested subsidy provided to students from low-income households. Hence, receipt of school financial assistance serves as a good proxy for household income.

Information regarding teacher characteristics such as gender, years of teaching experience within the prefecture, and rank (promotion status of teachers) are available.[10] We match this information about teachers with a classroom, and thus with students in the class, to which they are assigned in a year at school.

For the estimation, we restrict our sample to students from 2nd to 5th grades because we focus on the estimation of classroom teachers' total effect on students in primary schools. We exclude the first grade because the information of first grade students is not available before 2013, restricting sample size severely. We exclude sixth grade students from our analysis because test to measure educational achievement in 6th grade takes place in 7th grade in middle schools where test-taking environment is very different from ones in primary schools. As explained in Section 2, classroom teachers teach all subjects to students in their class for a year in primary schools. In contrast, in middle school in Japan, classroom teachers are assigned, but subject teachers teach each subject. Since we cannot identify subject teacher in middle schools in our dataset, we have restricted our sample only to the students in primary schools. We focus on the periods from 2010 to 2016 in our analysis due to data availability.

[Table 1]

Table 1-A summarizes the number of observations of the sample used for the TVA estimation. On average, in each year and grade, about 4000 students are observed. The total number of observations is 118,356. The bottom panel reports the number of teachers observed

[9] In the municipality that is the focus of our analysis, there are 69 public primary schools and 35 public middle schools in 2020. In this municipality, there is no private primary school, and there is one private middle school. In the estimation sample, we have 72 primary schools because 3 schools were merged with other schools after 2017.
[10] We will explain the rank system for teachers in Section 6.

in each year. We use about 600 teachers each year for the estimation. In total, 4197 teacher-year observations were used.

Descriptive statistics of the sample are reported in Table 1-B. The average and the standard deviation of test scores for language arts and mathematics in the estimation sample are close to 0 and 1, respectively, because these test scores are standardized by subject, grade, and year.[11] As a control of time-varying student characteristics, we include a dummy variable indicating receipt of school financial assistance. School financial assistance is provided to students to purchase school supplies, school lunches, school trip expenses, and medical fees if the municipal board of education acknowledges that students need such assistance because of low household income.[12] In our sample, 33.5% of students received school financial assistance from the municipality. As other controls at class level, we include female share and class size. Average female share is about 50%. Average class size is about 31 students per class. In our sample, we do not observe extremely small classes with less than 7 students in a class. The maximum class size is 40 students, which is the cap for all elementary schools in Japan. Year-school-class average of female share is about half, and the share of school financial assistance recipients is about 34% at class level and school level, respectively.

The bottom panel summarizes statistics of teacher characteristics. Teachers in our sample have 11 years of teaching experience on average. Regarding teacher promotion information, 28.6% of teachers are ranked as year-head teachers, who are leaders of teachers in a grade (this paper will refer to them as "chief teacher"). Another 6.5% of teachers are ranked as senior teachers, who serve as head of all classroom teachers and rank above a head teacher and hold the third-highest rank, following principal and vice-principals (call them "managing teacher"). The remainder are ordinal teachers.

5. Results of TVA Estimation

5.1 Results of Regression Analysis

The estimates of the coefficients of the regression equation explained in Section 3 are reported in Table 2.

[Table 2]

[11] The mean and standard deviation of standardized test scores are not exactly zero and one, respectively, in the estimation sample because test scores are standardized among all test takers and because some observations are excluded from the estimation sample due to lack of data availability (missing information of school financial assistance status).
[12] With high probability, students from households on welfare are offered school assistance.

The first six columns report the estimation results for the equation, with mathematics test scores as the dependent variable. The first column includes only those teacher-fixed effects necessary for the following TVA estimation. Receipt of school financial assistance has a negative and statistically significant coefficient. The share of students receiving school financial assistance is negatively associated with math test scores as well. Female students' math scores are lower by 0.04 standard deviation than male students. However, the share of female students per class and per school for a given year is positively associated with math test scores. The coefficient on the class size is negative and statistically significantly different from zero. The estimated coefficient indicated that the reduction of class size by one student increase the math test score by 0.005 standard deviation.

The second column reports the estimation results for the specification with school fixed effects. The coefficients on the individual level covariates are almost unchanged by inclusion of school fixed effects. The coefficients on school-year share of female students lost statistical precision by including school fixed effects because the school level share of female students has little variation over time. The third column is for the specification with year-school fixed effects. The results are similar to those in the second column.

The fourth to sixth columns report the estimation results with student fixed effects. The coefficient on school financial assistance is negative but insignificant. This observation of the results without student fixed effects indicates that it is not the receipt of school financial assistance itself but the long-lasting and time-invariant factors such as long-term household income and home environment matter for learning outcomes of mathematics. Class size is negatively associated with math test score in all but the last specification.

The seventh to twelfth columns report the estimation results for language arts test scores. Similar to the case for math test score, school financial assistance has a negative and statistically significant coefficient without controlling for student fixed effect. The share of students receiving school financial assistance is negatively associated with test scores. Contrary to the case for math test scores, female students' language arts score is higher than male students by 0.18 standard deviation. The share of female students in class and in school for a given year is also positively associated with language arts test scores. The coefficient on the class size is negative and statistically significantly different from zero. The estimated coefficient indicates that the reduction of class size by one student is associated with an increase of math test score by 0.003 standard deviation.

The eighth column with school fixed effects and the ninth column with school-year fixed effects demonstrate very similar to the results in the seventh column. The coefficients on the individual level covariates are almost unchanged by inclusion of school or school-year fixed effects. The tenth through twelfth columns report the estimation results with student fixed effects. Similar to the case for mathematics, the coefficient on school financial assistance is negative but insignificant. Class size is negatively associated with language arts test score in all specifications.

In a nutshell, we found that female students were better in language arts and worse in mathematics than male students. We found that small class size was associated with higher test scores of both subjects. Finally, school financial assistance was negatively associated with test scores, but this association disappeared when student fixed effects were included.

5.2 Teacher Value-Added

Teacher value-added are estimated by the method explained in Section 3. We estimate the TVA using test score residuals with the specification-controlling school-year fixed effect without student fixed effects (the results reported in the third column for mathematics and the ninth column for language arts in Table 2). This is for the sake of comparison to the estimates of teacher effectiveness using value-added formulation without student fixed effects, reported in the literature such as Chetty, Friedman and Rockoff (2014a).

[Table 3]

Table 3 summarizes the TVA model parameters' estimates. Panel A reports autocovariance and autocorrelation vectors for language arts and mathematics test scores, respectively. All autocovariances are statistically significantly different from zero for both language arts and mathematics. More importantly, autocorrelations are high and declined slowly. For math test score, autocorrelations are larger than 0.5 up to the fourth lag and remain at above 0.4 for the rest. For language arts test scores, autocorrelations are slightly lower than those for math test scores, and they remain approximately 0.45 up to the fifth lag.

Using these results, we estimate the standard deviation of TVA. We provide two estimates for this. The first estimate is based on the first-order autocovariance as the lower bound of within-year TVA variance. Panel B of Table 3 reports the estimates of teacher standard deviation based on the covariance between TVA and its first lag. The lower bound of the standard deviation of TVA is 0.150 for language arts and 0.174 for mathematics. Similar to the U.S. case reported in Chetty, Friedman, and Rockoff (2014a), we find a larger standard deviation for math test scores than for language arts test scores.

The second estimate uses a quadratic regression equation to predict the contemporaneous variance from autocovariances. Specifically, we regress the log of the first six autocovariances in Panel A on the time lag and time lag squared and extrapolating to zero from the estimate of the constant term to infer the within-year covariance.[13] The estimates based on the second method provide very similar but slightly larger estimates of TVA standard deviation. The standard deviation of TVA is 0.155 for language arts and 0.179 for mathematics. Our estimates of the standard deviation of TVA are close to the ones reported in Chetty, Friedman, and Rockoff (2014a), but our estimates are 25% higher than those in the U.S. for language arts (i.e., English), and 10% higher for mathematics. These findings indicate that the variations of TVA in Japan have magnitudes comparable to or even larger than those in the U.S.

5.3 Prediction Unbiasedness and Student-Teacher Matching

As stressed in Chetty, Friedman, and Rockoff (2014a), the estimated TVA is the best linear predictor of the test score residuals on theoretical ground. To confirm this point empirically, we regress residual test score on the estimated TVA and see how well the estimated TVA predicted the observed test score residuals for students.

[Table 4]

The first and second columns report coefficients from OLS regression in which we regress residual test scores on the estimated TVA. We can observe that the coefficients on the estimated TVA are close to those we expected. From these results, we can confirm that the estimated TVA using the best linear predictor of the test score residuals serve as good predictors of annual gains of student test scores.

[Figure 1]

To confirm this point visually, Panel A of Figure 1 plots the relationship between the actual test score residual and TVA for language arts and mathematics, respectively. The slope of the estimated line is almost 45-degrees and thus we can again confirm that our estimates of TVA predicted average actual annual gain of teachers' test scores.

[13] Another way to estimate the standard deviation is to estimate it using multiple observations of classes taught by the same teacher in one year, as applied to a middle school case in Chetty, Friedman, and Rockoff (2014a). We cannot apply this method with our sample because teachers in primary schools teach only one class per year.

One potential threat to TVA estimation is endogenous matching between students and teachers. If the sorting of students across teachers is due to time-invariant unobservable student characteristics or previous test scores (static sorting in Kane and Staiger, 2008), the value-added formulation controls the effect of sorting in the estimation of TVA by removing student fixed effect or previous test scores. However, if the sorting is based on the growth of test scores (that is, dynamic sorting in Kane and Staiger 2008), it may cause biases in the estimation of TVA. If the dynamic sorting causes serious bias in the estimation of TVA, we should observe strong correlation between the growth of past residual test scores and future TVA.

The third and fourth columns in Table 4 report the OLS coefficient of regression of the predicted test score residual using test scores from t-2 on the estimated TVA in year t for language arts and mathematics, respectively. The estimated coefficients are close to zero and insignificant at a conventional level of significance. These estimates indicate that past test score and past test score growth are not significantly correlated to the estimated TVA. Hence, we may conclude that potential bias caused by dynamic sorting is not serious in our data and thus in the estimates of TVA. Similar to the reason proposed by Chetty, Friedman, and Rockoff (2014a), inclusion of past test scores in a flexible way controls the potential determinants of sorting and thus the estimated value-added based on test score residuals are less affected by potential student-teacher sorting in our sample.

6. Teacher Value-Added and Teachers' Observable Characteristics

In this section, we examine which observable characteristics of teachers are associated to TVA by regressing estimated TVA on gender, years of experience, and age of teachers. Then we examine whether the estimated TVA is associated with the probability of teacher promotion to a higher rank.

6.1 Gender, Teaching Experience, and Age

In Section 5, we estimated TVA. As we allowed drift of TVA over a year, we obtained 2674 observations of the TVA of 974 individual teachers. In the estimation of the correlations between TVA and gender, years of experience, and teacher age, it is important to consider the fact that TVA is highly correlated between subjects taught by the same teacher because both subjects are taught by the same classroom teacher in public primary schools in Japan. In fact, the raw within-teacher correlation coefficient of the 2674 estimated TVA between language arts and mathematics is 0.6578, and the correlation of the same-teacher average of TVA estimates between the subjects was 0.6735 for 974 individual teachers. To take into account this high correlation of TVA between subjects, we estimate the determinants of TVA for both mathematics and language arts simultaneously by the seemingly unrelated regression.

[Table 5]

Table 5 reports the results of the regression analysis. The first two columns of Table 5 report the estimates of the specification including only teacher's gender and years of teaching experience in the prefecture. We report two standard errors: standard errors without clustering are in parentheses, and those with clustering at school level are in brackets. The coefficient on the female dummy is positive and statistically significant at least at the 10% level for language arts. Contrary, the coefficient on the female dummy is negative but insignificant for mathematics. The coefficients on the years of teaching experience are positive and statistically significant for both language arts and mathematics without clustering at the school level.

The second two columns report the estimates of the specification including only teachers' gender and age. The coefficient on the female dummy is positive and statistically significant at least to the 5% level for language arts, and it is negative and insignificant for mathematics. The coefficients on teachers' age are positive but insignificant for both language arts and mathematics.

Finally, the third two columns report the estimates of the specification including both years of teaching experience and teacher age on top of teacher gender. Similar to the first two sets of results, the coefficient on the female dummy is positive and statistically significant for language arts, but it is insignificant for mathematics. More interestingly, the coefficients of the years of teaching experience are positive and statistically significant for both language arts and mathematics. The magnitude of the coefficients are more than four-times larger than those in the first result. In fact, the coefficients on teacher age are negative and statistically significant. Because years of teaching experience and age of teacher are highly correlated with each other, it is important for isolating the association between teaching experience and TVA to control for teacher age. Hence, these results imply that teaching experience is positively associated with TVA conditional on teacher's age.

One interesting question concerns how the estimated TVA correlated between subjects. The correlation coefficients of the residuals are reported in Table 5 for each regression model. The smallest correlation coefficient is 0.6557, which is 0.3% smaller than raw correlation coefficient. This means that, although the observable teacher characteristics common to both subjects could account for the variation of TVA to some extent, it explains only small fraction of the cross-subject correlation of TVA. This is not surprising with the value of adjusted R-squared less than 2%.

In summary, female teachers had higher value-added for language arts, but no difference in TVA for mathematics between male and female teachers was found. Conditional on teacher age, teaching experience was positively associated with TVA for both subjects. These results indicate that, although it remains difficult to find a good teacher based on observable teaching characteristics due to the low explanatory power of these observable characteristics, these observable characteristics nevertheless inform TVA.

6.2 Heterogenous Relation by Decile of Teacher Value-Added

The results reported in Table 5 summarize the statistical associations between the estimated TVA and observable characteristics of teachers on average. The relationship between teachers' observable characteristics and their value-added measurements may be heterogenous along the distribution of the estimated TVA. To explore this heterogenous relationship, we estimate quantile regression models of TVA by decile for each subject. We use the specification of the regression models reported in the third column with both years of teaching experience and teacher's age in Table 5. We estimate these equations for each subject separately. Standard errors are clustered at school level.

[Figure 2]

Figure 2 summarizes the estimated coefficients, indicated with a real line and the 95% confidence interval with a dashed line (full estimation results are reported in Table A1). Panel A depicts the coefficients on the female dummy. For language arts, the coefficients are positive, almost uniformly distributed, and marginally significant for most of the ranges of estimated TVA. For mathematics, however, the coefficients are close to zero and insignificant for all deciles. These results indicate that female teachers are better in language arts for almost all ranges of the TVA distribution.

Panel B describes the coefficients for the years of teaching experience conditional on teachers' gender and age. For language arts, the coefficients are always positive, almost uniformly distributed, and significant for the most of ranges of estimated TVA. For mathematics, the coefficients are positive but statistically significant only for the lower side of TVA distribution. These results indicate that teaching experience is important for all ranges of TVA distribution for language arts, but it is important only for teachers with relatively low value-added for mathematics. For teachers with relatively low TVA, accumulating teaching experience could be a solution to improve TVA, but the finding does not explain the high value-added of teachers in mathematics.

Panel C describes the coefficients on the age of teachers, conditional on teachers' gender and experience. The results are almost mirror images of those for teaching experience: for language arts, the coefficients are negative, almost uniformly distributed, and significant for the most of ranges of estimated TVA. For mathematics, the coefficients are positive but statistically significant only for the lower side of TVA distribution.

In summary, teaching experience and teacher age were associated with TVA for both language arts and mathematics, and female teachers tend to have higher value-added measurements in language arts than male teachers. For language arts, these observable characteristics are uniformly important for TVA. For mathematics, teaching experience and age matter for teachers with relatively low TVA values.

6.3 Teacher Promotion

A potential use of the estimated TVA is to evaluate teachers' performance based on these measures. For example, Staiger and Rockoff (2010), Chetty, Friedman, and Rockoff (2014b), and Bau and Das (2020) discuss extensively potential policies using the estimates of TVA to recruit and retain high performing teachers and to dismiss low-performing ones.

In the institutional setting where we derived our data, TVA measures have not been calculated or used for teacher evaluation. However, it may be possible for members of the board of education and school principals to employ TVA or a proxy for it at least in evaluations for teacher promotion. Hence, it is interesting and relevant to explore whether teacher quality measured by TVA in mathematics and language arts test scores is associated with the promotion of teachers.

In our sample, there are three ranks for teachers: regular teachers, chief teachers, and managing teachers. The role of managing teacher, one rank below vice principal, is to assist the principal and vice-principals, coordinate between the principals and senior teachers, and lead the senior teachers in maintaining the school environment. The main role of chief teachers, one rank below managing teachers, is to support managing teachers and teach and supervise regular teachers and younger school staff members. Because both managing teachers and chief teachers have their own class to teach, we can estimate TVA for these teachers with the same method used to estimate TVA of regular teachers.

To earn promotion from regular to chief teacher, a teacher's age must be above 30 with more than 8 years of teaching experience. Hence, it is natural to expect that both teaching experience and age would be positively correlated to promotion probability to chief teacher. More importantly, to be a chief teacher, a teacher must pass the essay exam and have a good evaluation from the school principal. Hence, if school principals could identify teachers who produce significant standardized achievement gains, as found in Jacob and Lefgren (2008), and could then evaluate teachers based on teaching performance, we would expect to observe positive correlation between TVA and promotion to chief teacher.

For promotion to and certification as managing teachers, there are two criteria: Criterion A is for teachers under 44 years old; Criterion B is for teachers between 46 and 53 years of age. In Criterion A, applicants need to pass the essay exam and produce a research paper on pedagogy, and they are interviewed by a committee of the prefectural board of education. In Criterion B, applicants need to pass the same exam as in Criterion A, excluding the writing of a research paper on pedagogy. Because of the presence of these two sets of criteria for promotion to managing teacher with different eligibility requirements in terms of age, the correlation between age and probability of promotion to managing teachers may be less obvious than in those cases for chief teachers.

To examine the probability of promotion and its relation to TVA, we estimate a linear probability model with a dummy variable indicating promotions to chief teacher and managing teacher as dependent variables, respectively. In the equation for the promotion to chief teacher, we exclude samples of managing teachers because managing teachers would not be demoted to a lower rank. We control for grade fixed effects, survey year fixed effects, and school fixed effects. Standard errors are clustered by teacher level.

[Table 6]

Table 6 reports the results of the estimations of the probability for promotion to chief teacher (the first three columns) and to managing teachers (the last three columns). The specification in the first column includes only math TVA in addition to other controls. The coefficient is positive but insignificant. The second column shows the model with TVA for language arts. The coefficient is positive and statistically significant. The magnitude of the coefficient is sizable: improvement of TVA by one standard deviation (i.e., 0.15) is associated with an increase of the probability of promotion to a chief teacher by 6.7 percentage points. The third column shows the model with both language arts and math TVA. The findings show that only the TVA for language arts is positively associated with the promotion to a chief teacher. This result indicates the possibility that the board of education members and school principals have some sense of and use for TVA in teacher evaluation for promotion to chief teacher even without formal measurement of TVA.

It is noteworthy to mention the coefficients for the other control variables in the model for the promotion to chief teacher. The coefficients on female dummy variable are always positive but statistically insignificant. Teaching experience has positive and statistically significant coefficients. Similarly, the coefficients on age are positive and statistically significant. These results are consistent with the institutional setting that age and teaching experience are parts of requirements to be eligible to apply for the promotion to a chief teacher.

The fourth to sixth columns report the estimation results of the probability to managing teachers. Unlike the results for chief teachers, none of the coefficients on TVA is statistically significant. This may be because several skills other than teaching, such as management and administrative knowledge, are required and evaluated more in the promotion process to managing teachers. Although these coefficients are insignificant, it is interesting to observe that the coefficient for math TVA tended to be larger than that for language arts, in contrast to the case for chief teachers.

On the other controls, the coefficients on the female dummy variable are negative and statistically significant, implying gender imbalance exists in determining promotion to the level of managing teacher. Teaching experience has positive and statistically significant coefficients, reflecting the fact that one of the eligibility requirements to be a managing teacher is to have teaching experience as a chief teacher. Age is not statistically significantly associated with promotion to the managing teacher level. As we have discussed above, this may be because while criterion B sets minimum age, criterion A sets the maximum age for eligibility.

7. Conclusion

In this study, we estimated TVA using language arts and mathematics test scores of students in public primary schools in Japan. Applying the method proposed by Chetty, Friedman, and Rockoff (2014a) to administrative panel data from a large local municipality of Japan, with additional control of school-year fixed effects, we found that the distribution of TVA has a positive variance for both language arts and mathematics subjects, and that their size is comparable to those observed in the United States. Our estimates provide the first estimates of TVA in public primary schools in Japan, and they show the robustness of TVA estimates in an advanced country with a very different educational institutional setting.

Using the TVA estimates, we examined their associations with observable teacher characteristics such as gender, teaching experience, and age as well as to teacher promotion. We found that those observable characteristics of teachers are statistically significantly associated with the estimated TVA, and that the estimated TVA is positively associated with the promotion of teachers. These findings suggest that even in a country without explicit measurement of TVA, school principals identify teacher effectiveness when measured by standardized academic achievement student gains, and they use it implicitly to determine teacher promotion in their schools. Study of the effects of disclosure of TVA in strengthening accountability of schools and teachers on students' short- and long-term performances would be a valuable topic for future research.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(1):95–135.

Azam, Mehtabul, and Geeta Gandhi Kingdon. 2015. "Assessing Teacher Quality in India." *Journal of Development Economics* 117(C):74–83.

Bacher-Hicks, Andrew, Thomas J. Kane, and Douglas O. Staiger. 2014. "Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles." NBER Working Papers 20657.

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

Buhl-Wiggers, Julie, Jason T. Kerwin, Jeffrey A. Smith, and Rebecca Thornton. 2019. "Teacher Effectiveness in Africa: Longitudinal and Causal Estimates." Presented at the Asian and Australasian Society of Labour Economics 2019 Singapore conference.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104(9):2593–2632.

__________. 2014b. "Measuring the Impacts of Teachers I: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104(9):2633–2679.

__________. 2016. "Using Lagged Outcomes to Evaluate Bias in Value-Added Models." *American Economic Review: Papers and Proceedings* 106(5):393–399.

Harris, Douglas N., and Tim R. Sass. 2011. “Teacher Training, Teacher Quality and Student Achievement.” *Journal of Public Economics* 95:798–812.

Hanushek, Eric A. 1971. “Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data.” *American Economic Review* 60(2):280–288.

__________. 1992. “The Trade-off Between Child Quantity and Quality.” *Journal of Political Economy* 100(1):84–117.

Hanushek, Eric A., John. F. Kain, Daniel M. O’Brien, and Steven G. Rivkin. 2005. “The Market for Teacher Quality.” Working Paper 11154. National Bureau of Economic Research.

Hanushek, Eric A. 2006. “School Resources.” *Handbook of the Economics of Education* (2):865–908.

Hanushek, Eric A., and Steven G. Rivkin. 2006. “Teacher Quality.” *Handbook of the Economics of Education* 2:1051–1078.

Hanushek, Eric A., and Steven G. Rivkin. 2010. “Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools?” National Bureau of Economic Research Working Paper 15816.

Hojo, Masakazu. 2012. “Determinants of Academic Performance in Japan.” *Japanese Economy* 39:3–29.

Hojo, Masakazu, and Takashi Oshio. 2012. “What Factors Determine Student Performance in East Asia? New Evidence from the 2007 Trends in International Mathematics and Science Study.” *Asian Economic Journal* 26:333–357.

Jacob, Brian A., and Lars Lefgren. 2008. “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics* 26(1):101–136.

Jackson, Kirabo. 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, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. “What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City.” *Economics of Education Review* (27):615–631.

Kane, Thomas J., and Douglas O. Staiger. 2008. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” National Bureau of Economic Research Working Paper 14607.

Koedel, Cory, and Julian 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.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. “How Large Are Teacher Effects?” *Educational Evaluation and Policy Analysis* 26(3):237–257.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. “Teachers, Schools, and Academic Achievement.” *Econometrica* 73(2):417–458.

Rockoff, Jonah E. 2004. “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review* 94(2):247–252.

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

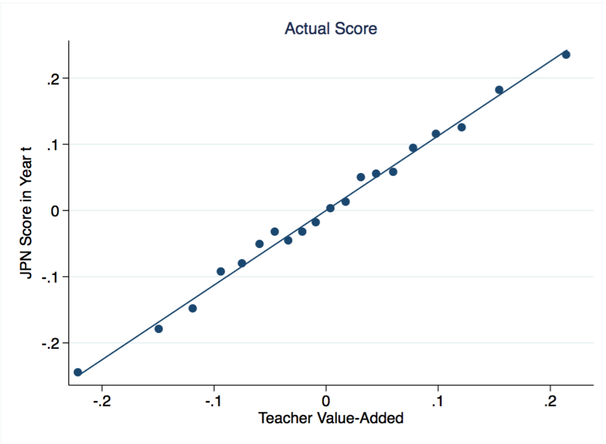
Staiger, Douglas O., and Jonah E. Rockoff. 2010. “Searching for Effective Teachers with Imperfect Information.” *The Journal of Economic Perspectives* 24(3):97–117.

Tanaka, Ryuichi, and Kazumi Ishizaki. 2018. “Do Teaching Practices Matter for Students’ Academic Achievement? A Case of Linguistic Activity.” *Journal of the Japanese and International Economies* 50:26–36.

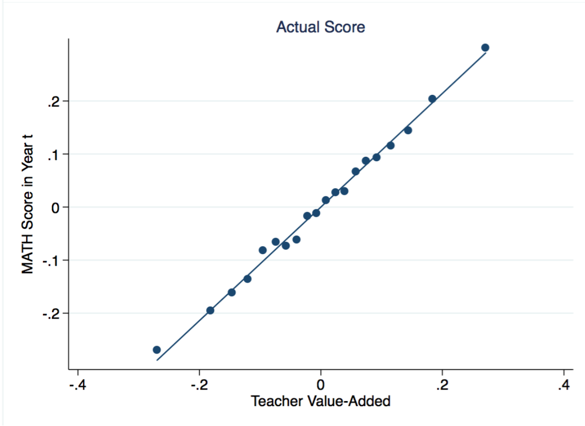
Figure 1: Effects of Teacher Value-Added on Actual and Predicted Scores

Panel A: Actual Score

Japanese Language Art

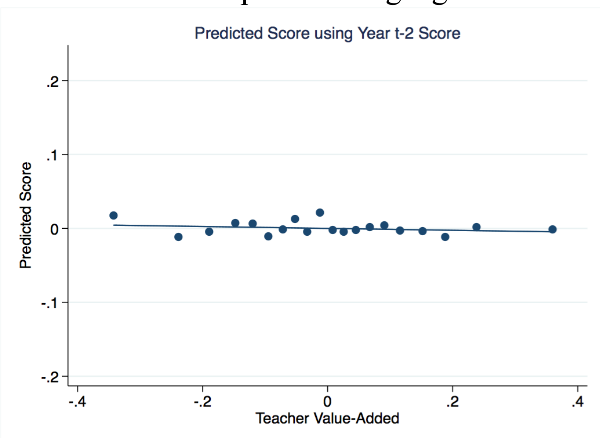


Math



Panel B: Predicted Score using year t-2 score

Japanese Language Art



Math

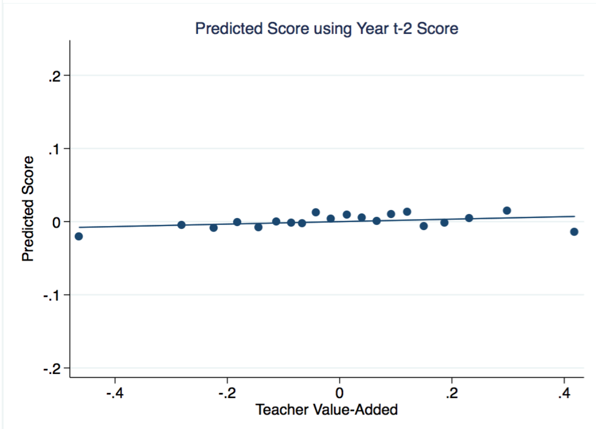
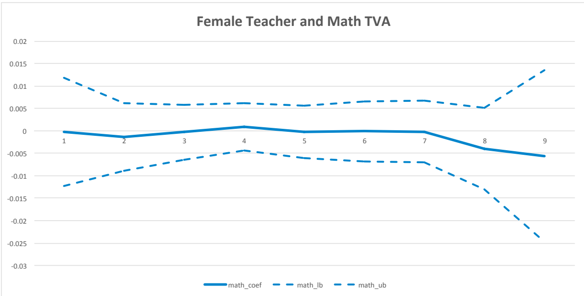
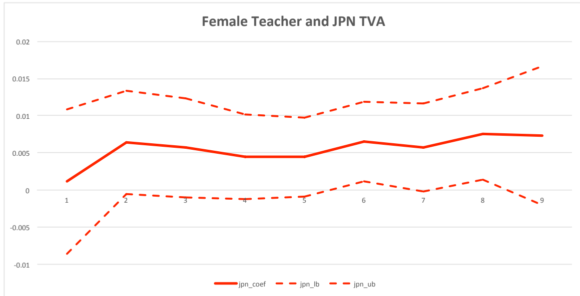


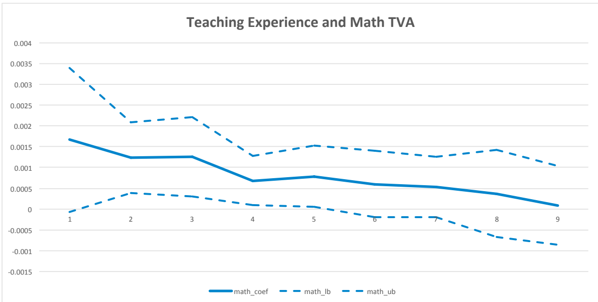
Figure 2: Teachers' Characteristics and Teacher Value-Added by Decile

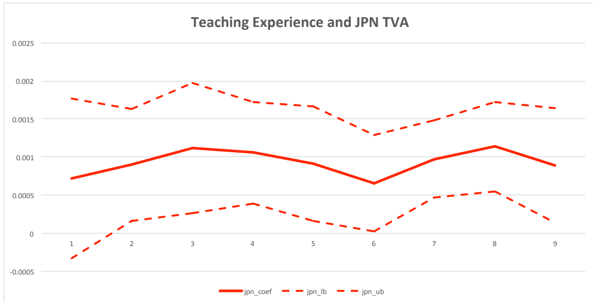
A. Female Teacher



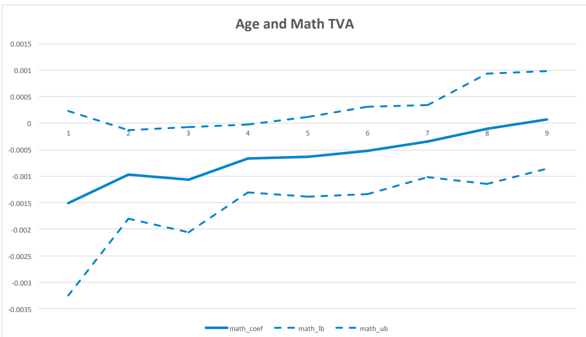


B. Teaching Experience





C. Age



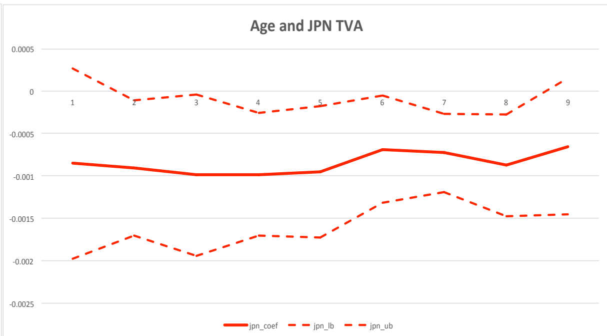


Table 1-A: Number of Student and Teacher Observations

Year	2010	2011	2012	2013	2014	2015	2016	Total
Students								
Grade								
2	4,118	4,228	4,259	4,092	4,556	4,111	3,786	29,150
3	4,181	4,166	4,362	4,256	4,283	4,116	3,875	29,239
4	3,804	4,130	4,640	4,251	4,335	4,287	3,773	29,220
5	3,996	3,946	5,019	4,684	4,678	4,685	3,739	30,747
Total	16,099	16,470	18,280	17,283	17,852	17,199	15,173	118,356
Teachers	603	589	630	603	625	602	545	4,197

Table 1-B: Descriptive Statistics

Variable	Mean	Std.Dev.	Min	Max
Students				
Language Arts score	0.007	0.988	-5.839	2.051
Mathematics score	0.006	0.988	-5.982	1.557
School financial assistance	0.335	0.472	0.000	1.000
Female	0.494	0.500	0.000	1.000
Class size	31.288	4.335	7.000	40.000
Female (class average)	0.491	0.066	0.182	0.718
Female (school average)	0.491	0.027	0.369	0.559
School financial assistance (class average)	0.345	0.101	0.111	0.650
School financial assistance (school average)	0.339	0.129	0.000	0.895
Teachers				
Years of teaching experience in prefecture	11.081	10.308	1.000	41.456
Female teacher	0.643	0.479	0.000	1.000
Chief teacher	0.286	0.430	0.000	1.000
Managing teacher	0.065	0.236	0.000	1.000

Table 2: Estimated Results of Determinants of Test Scores

	Mathematics						Language Arts					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
School financial assistance	-0.0851***	-0.0850***	-0.0847***	-0.0026	-0.0028	-0.0020	-0.0775***	-0.0775***	-0.0773***	-0.0160	-0.0156	-0.0167
	(0.0040)	(0.0040)	(0.0039)	(0.0102)	(0.0102)	(0.0101)	(0.0040)	(0.0040)	(0.0040)	(0.0106)	(0.0106)	(0.0105)
Female	-0.0393***	-0.0394***	-0.0393***				0.1815***	0.1815***	0.1814***			
	(0.0035)	(0.0035)	(0.0034)				(0.0035)	(0.0035)	(0.0035)			
Class size	-0.0050***	-0.0051***	-0.0034***	-0.0049***	-0.0054***	-0.0025	-0.0028***	-0.0032***	-0.0020**	-0.0024*	-0.0030**	-0.0036**
	(0.0007)	(0.0008)	(0.0009)	(0.0013)	(0.0013)	(0.0017)	(0.0007)	(0.0008)	(0.0009)	(0.0014)	(0.0014)	(0.0017)
Female (class average)	0.0793*	0.0647	0.0132	-0.1453	-0.1916	-0.2722*	0.0889**	0.1013**	0.1105**	0.2971**	0.2215*	0.0418
	(0.0439)	(0.0449)	(0.0480)	(0.1254)	(0.1254)	(0.1461)	(0.0447)	(0.0455)	(0.0490)	(0.1306)	(0.1314)	(0.1523)
School financial assistance (class average)	-0.0533	-0.0287	-0.0122	-0.0384	-0.0513	0.0327	-0.0121	0.0155	0.0028	0.0145	0.0111	0.0120
	(0.0331)	(0.0338)	(0.0362)	(0.0449)	(0.0451)	(0.0503)	(0.0336)	(0.0343)	(0.0370)	(0.0470)	(0.0471)	(0.0532)
Female (school average)	0.4345***	0.2151	0.0623	-0.0923	-0.0752	-3.4888	0.6415***	0.3680**	46.8349***	0.4360**	0.3376	45.2240***
	(0.1406)	(0.1568)	(12.1152)	(0.1915)	(0.1947)	(11.4146)	(0.1458)	(0.1614)	(11.7717)	(0.2062)	(0.2086)	(12.1983)
School financial assistance (school average)	-0.2308***	-0.4267***	-0.0736	-0.2936**	-0.5314***	2.2504	-0.5524***	-0.5754***	-4.9748***	-0.1420	-0.3147*	-1.8093
	(0.0761)	(0.1249)	(1.2532)	(0.1319)	(0.1492)	(1.4747)	(0.0784)	(0.1277)	(1.2479)	(0.1398)	(0.1611)	(1.6287)
Constant	0.0423	0.1194	0.0684	0.4011*	0.2046	1.2530	0.0080	-0.0331	-22.3942***	-0.0192	0.0624	-20.9655***
	(0.1864)	(0.2419)	(5.8601)	(0.2331)	(0.3125)	(5.3539)	(0.1348)	(0.2068)	(5.6919)	(0.1988)	(0.2781)	(5.7382)
Student fixed effects	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes
Teacher fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
School fixed effects	no	yes	no	no	yes	no	no	yes	no	no	yes	no
School-year fixed effects	no	no	yes	no	no	yes	no	no	yes	no	no	yes
Observations	118356	118356	118356	118356	118356	118356	118356	118356	118356	118356	118356	118356
r2	0.6171	0.6179	0.6265	0.1431	0.1446	0.1693	0.6013	0.6020	0.6089	0.1455	0.1477	0.1686
r2_adjusted	0.6099	0.6105	0.6178	0.1280	0.1289	0.1512	0.5938	0.5942	0.5999	0.1303	0.1320	0.1504

Standard errors in parentheses. * indicates 10 percent, ** for 5 percent, *** 1 percent level of significance.

Table 3: Teacher Value-Added Model Parameters Estimates

	Mathematics	Language Arts
Panel A: Autocovariance and autocorrelation vectors		
Lag 1	0.030	0.023
obs. = 1422	(0.001)	(0.001)
	[0.575]	[0.490]
Lag 2	0.027	0.021
obs. = 997	(0.001)	(0.001)
	[0.529]	[0.483]
Lag 3	0.025	0.020
obs. = 592	(0.002)	(0.002)
	[0.529]	[0.487]
Lag 4	0.023	0.018
obs. = 332	(0.002)	(0.002)
	[0.505]	[0.468]
Lag 5	0.020	0.017
obs. = 150	(0.003)	(0.003)
	[0.480]	[0.459]
Lag 6	0.016	0.026
obs. = 49	(0.003)	(0.005)
	[0.414]	[0.642]
Panel B: Within-year variance components		
Total SD	0.642	0.651
Individual-level SD	0.607	0.621
Class+teacher level SD	0.209	0.195
Estimates of teacher SD		
Lower bound based on lag 1	0.174	0.150
Quadratic estimates	0.179	0.155

Panel A reports the estimated autocovariance, the standard error of that covariance estimate clustered at the teacher level (in parentheses), and the autocorrelation (in brackets) of average test score residuals between classrooms taught by the same teacher. We measured these statistics at time lags ranging from one to six years (i.e., two classrooms taught six years apart), weighted by the sum of the relevant pair of class sizes. Each covariance is estimated separately for Language Arts and Mathematics and for elementary school classrooms. Panel B reports the raw standard deviation of test score residuals and decomposes this variation into components driven by idiosyncratic student-level variation, classroom shocks, and teacher-level variation. Row 1 provides a sum of the variances in rows 2 and 3 of panel B; row 3 provides a sum of the variances in rows 4 and 5. In elementary schools, we could not separately identify class-level and teacher-level standard deviations because we observed only one classroom per year. We used the square root of the autocovariance across classrooms at a one-year lag to estimate a lower boundary for the within-year standard deviation for elementary schools. We also report an estimate of the standard deviation by regressing the log of the first five autocovariances in panel A on the time lag and time lag squared and extrapolating to zero to estimate the within-year covariance.

Table 4: Estimates of Forecast Bias Using Lagged Score

	Score in year t		Pred. score using year t-2 score	
	Mathematics	Language Arts	Mathematics	Language Arts
Teacher VA	1.071	1.128	0.017	-0.013
	(0.031)	(0.036)	(0.009)	(0.013)
Observations	91,400	91,323	60,037	59,933

Each column reports coefficients from an OLS regression, with standard errors clustered by school-cohort in parentheses. The regressions are run on the sample used to estimate the baseline VA model, restricted to observations with a non-missing leave-out teacher VA estimate. There is one observation for each student-subject-school year in all regressions. Teacher VA is scaled in units of student test score standard deviations and is estimated using data from classes taught by the same teacher in other years. Teacher VA is estimated using the baseline control vector, which includes: a cubic in lagged own- and cross-subject scores, interacted with the student's grade level; student-level characteristics including gender and receipt of school financial assistance; class size; cubics in class and school-grade means of lagged own- and cross-subject scores, interacted with grade level; class and school-year means of all the student-level characteristics; and grade dummy and year-school dummies. In columns 1 and 2, the dependent variable is the student's test score in a given year and subject. In columns 3 and 4, the dependent variable is the predicted value generated in the same way from twice-lagged test scores.

Table 5: Teacher Value-Added and Teachers' Observable Characteristics

	(1)		(2)		(3)	
	Math	Language	Math	Language	Math	Language
Female	-0.0012	0.0053	-0.0007	0.0058	-0.0015	0.0049
	(0.0024)	(0.0021) **	(0.0024)	(0.0021) ***	(0.0024)	(0.0021) **
	[0.0032]	[0.0028] *	[0.0032]	[0.0028] **	[0.0032]	[0.0028] *
Teaching experience	0.0002	0.0002			0.0008	0.0010
	(0.0001) **	(0.0001) **			(0.0003) ***	(0.0002) ***
	[0.0002]	[0.0001]			[0.0004] **	[0.0003] ***
Age			0.0001	0.0000	-0.0006	-0.0009
			(0.0001)	(0.0001)	(0.0003) **	(0.0002) ***
			[0.0002]	[0.0002]	[0.0004] *	[0.0003] **
Correlation of residuals between subjects						
Raw = 0.6578		0.6566		0.6571		0.6557
Observations	2674	2674	2674	2674	2674	2674
r2 adjusted	0.0065	0.0087	0.0052	0.0072	0.0082	0.0134

All specifications include grade fixed effects. All coefficients are estimated by seemingly unrelated regression. Standard errors without and with clustering at school level are reported in parentheses and brackets, respectively. * 1%. ** 5%, *** 1% significance of coefficients.

Table 6: Teacher Value-Added and Teachers' Rank

	Chief Teacher			Managing Teacher		
TVA_math	0.2010		-0.0693	0.1246		0.1116
	(0.1522)		(0.2004)	(0.1269)		(0.1438)
TVA_language		0.4507 **	0.5030 **		0.1091	0.0240
		(0.1814)	(0.2398)		(0.1340)	(0.1454)
Female	0.0211	0.0189	0.0185	-0.0933 ***	-0.0940 ***	-0.0934 ***
	(0.0218)	(0.0219)	(0.0219)	(0.0175)	(0.0174)	(0.0175)
Teaching experience	0.0268 ***	0.0266 ***	0.0267 ***	0.0060 ***	0.0060 ***	0.0059 ***
	(0.0023)	(0.0023)	(0.0023)	(0.0014)	(0.0014)	(0.0014)
Age	0.0072 ***	0.0075 ***	0.0074 ***	0.0015	0.0015	0.0015
	(0.0023)	(0.0023)	(0.0023)	(0.0012)	(0.0012)	(0.0012)
Observations	2503	2503	2503	2674	2674	2674
r2_adjusted	0.5462	0.5475	0.5473	0.1459	0.1455	0.1456

All specifications include grade fixed effects, survey year fixed effects, school fixed effects, and constant term. Standard errors clustered by teacher_ID are in parentheses. * 1%. ** 5%, *** 1% significance.

Table A1: Teacher Value-Added and Teachers' Observable Characteristics by Decile

Panel A: Mathematics	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female teacher	-0.0002	-0.0013	-0.0003	0.0010	-0.0002	-0.0001	-0.0002	-0.0040	-0.0056
	(0.0062)	(0.0038)	(0.0031)	(0.0027)	(0.0030)	(0.0034)	(0.0035)	(0.0046)	(0.0098)
Teaching experience	0.0017*	0.0012***	0.0013**	0.0007**	0.0008**	0.0006	0.0005	0.0004	0.0001
	(0.0009)	(0.0004)	(0.0005)	(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)
Age of teacher	-0.0015*	-0.0010**	-0.0011**	-0.0007**	-0.0006*	-0.0005	-0.0003	-0.0001	0.0001
	(0.0009)	(0.0004)	(0.0005)	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0005)	(0.0005)
Observations	2674	2674	2674	2674	2674	2674	2674	2674	2674
R2	0.0083	0.0100	0.0084	0.0086	0.0097	0.0099	0.0097	0.0082	0.0067

Panel B: Language Arts	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female teacher	0.0011	0.0064*	0.0057*	0.0045	0.0044	0.0065**	0.0057*	0.0076**	0.0073
	(0.0049)	(0.0035)	(0.0034)	(0.0029)	(0.0027)	(0.0027)	(0.0030)	(0.0031)	(0.0048)
Teaching experience	0.0007	0.0009**	0.0011**	0.0011***	0.0009**	0.0007**	0.0010***	0.0011***	0.0009**
	(0.0005)	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Age of teacher	-0.0009	-0.0009**	-0.0010**	-0.0010***	-0.0009**	-0.0007**	-0.0007***	-0.0009***	-0.0006
	(0.0006)	(0.0004)	(0.0005)	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0003)	(0.0004)
Observations	2674	2674	2674	2674	2674	2674	2674	2674	2674
R2	0.0109	0.0149	0.0138	0.0150	0.0125	0.0138	0.0148	0.0147	0.0133

All specifications include grade fixed effects. Standard errors clustered at school level are in parentheses. * 1%. ** 5%, *** 1% significance.