The Australian Journal of Agricultural and Resource Economics

AARES AUSTRALASIAN AGRICULTURAL & RESOURCE ECONOMICS SOCIETY

Australian Journal of Agricultural and Resource Economics, 64, pp. 171–188

Incentivising teachers? Evaluating the incentive effect of China's teacher performance-based compensation reform in rural China*

Jian Zhang, Songqing Jin and Wei Si†

In this paper, we evaluate the incentive role of a teacher performance-based compensation reform in rural China. Using the value-added model widely adopted in the education literature, we first estimated the teacher effects on student academic scores with panel data of a large number of students and teachers from rural and urban schools in one county in a south-western province of China. The estimated teachers' value-add was then allowed us to examine the effectiveness of the 2009 teachers' compensation reform. We find that despite the strong intent of the performance-based compensation reform to improve student's academic performance, teachers' compensations are not closely tied to teachers' value-add to student academic achievement. This suggests that the performance-based compensation reform is not able to provide strong incentives for teachers to raise students' test scores and points towards the possible problems with the design and/or implementation of the reform.

Key words: China, performance-based compensation reform, teachers, value-added model.

1. Introduction

A large body of literature relating teacher quality to student academic performances has consistently shown that teacher quality has a significant effect on student achievement. Studies using large administrative datasets have also shown the existence of a large variation in teacher quality in terms of the teachers' value-addition to student academic achievements (Nye *et al.* 2004; Rockoff 2004; Rivkin *et al.* 2005; Aaronson *et al.* 2007; Kane and Staiger 2008; Kane *et al.* 2008; Jacob and Lefgren, 2008; Koedel and Betts 2009; Hanushek and Rivkin 2010; Rothstein 2010). This indicates teachers' qualifications, experience and motivations are critical to the success of students' learning.

*Songqing Jin is grateful for the financial support from Michigan State University's AgBioResearch, and the National Natural Science Foundation of China (71773110 and 71861147002).

†Jian Zhang is a Professor in School of Economics, Central University of Finance and Economics, Beijing, China. Songqing Jin (email: jins@msu.edu) is an Adjunct Professor at the China Academy for Rural Development, Zhejiang University, Hangzhou, China. Songqing Jin is also an Associate Professor at the Department of Agricultural, Food and Resource Economics, Michigan State University, East Lansing, Michigan, USA. Wei Si is a Professor at the College of Economics and Management, China Agricultural University, Beijing, China.

doi: 10.1111/1467-8489.12335

According to an official source, there are 109 million students and seven million teachers in elementary, middle and high schools in rural China (China Education Statistical Yearbook 2013). With an intention of incentivising teachers in improve their teaching performance in order to help the 109 million rural students' to achieve better academic outcomes, China's Ministry of Education initiated a performance-based compensation reform for elementary and middle school teachers. The overall goal of this reform initiative was to incentivise teachers for better performance by restructuring teachers' compensation on the basis of their performance.

There have been a large number of studies in the literature linking teachers' incentive to students' performance. It has been widely documented that salaries associated with teacher characteristics such as education and experience are only weakly correlated with student academic achievement (Rivkin *et al*. 2005; Podgursky and Springer 2007; Hanushek and Rivkin 2010), and an unconditional increase in teacher salaries is unlikely to improve student learning (de Ree *et al*. 2015). In response, performance-based pay schemes have been increasingly used in both developed and developing countries to better align teacher compensation with students' achievements (Bruns *et al*. 2011; Neal 2011; Loyalka *et al*. 2019). In a broader context, the pay-for-performance payment schemes tied to results have been increasingly used to incentivise public servants to provide better public services (e.g. in the health service sector – Loevinsohn and Harding 2005; Bloom *et al*. 2006; Gertler and Vermeersch 2012). However, empirical evidence on the effectiveness of these performance-based pay schemes is rather mixed. While many studies have shown that the performance-based pay schemes have improved students' grades (Woessmann 2011; Loyalka *et al*. 2019), others failed to find any significant effect of these schemes on students' achievement (Fryer 2013; Barrera-Osorio and Raju 2015).

Whether a performance-based pay scheme actually incentivises teachers' performance and improves student learning depends crucially on its specific pay design (Bruns *et al*. 2011; Neal 2011) and implementation. For example, a recent study has shown that the pay-for-percentile design has a larger effect than other alternative designs (Loyalka *et al*. 2019). There is also evidence that in the case of a fixed amount of total resource, the net effect of a performance-based pay scheme can be ambiguous because promoting the targeted outcome (e.g. health) could crowd out resource that could be used to improve another outcome (e.g. education) (Sylvia *et al*. 2013). Taken all together, whether a performance-based scheme is effective or not is largely an empirical issue.

In this paper, we aim to evaluate whether the teacher performance-based compensation reform has indeed achieved its goal of incentivising teacher performance in rural China and if so, whether the teachers' incentive is translated into students' academic scores. To do so, we first estimated the teacher effect (the teachers' value-addition to student academic performance) using a large panel dataset containing information on students and teachers

from rural China, and the standard teachers' value-added model (VAM). One important feature of VAM is that we are able to estimate the teacher effect by holding other confounding factors constant including students' previous scores as well as individual characteristics and family background. To understand the effectiveness of the reform, we then examine how closely teachers' compensation is related to the teachers' value-addition to student academic performance after the teacher performance-based compensation reform.

It is important to note that ideally, to examine the incentive effects, we would want to compare teachers' behaviours and performances before and after the reform and draw conclusions about the causal effect of the performance-based reform on teachers' behavioural outcomes. Unfortunately, we do not have such data. Instead, we evaluate the effectiveness of the performance-based reform by relating a teacher's payment to the estimated teachers' value-addition to student test scores, a measure of a teacher's teaching effect on students' test scores. Our hypothesis is that if the reform is effective, then after a reform, a teacher with higher teacher effect should get higher payment than a teacher with lower teacher effect. This is an indirect way to test the effectiveness of the reform (i.e. the incentive effects of the reform).

Notwithstanding the good intent of the performance-based compensation reform to improve students' academic performance, this paper finds that teachers' compensation is not closely linked to the estimated teachers' value-addition to student test scores. First, our data show that after the reform, teachers' monthly compensation increases significantly, almost doubling from 1,589 to 3,108 yuan. The performance-based pay accounts for 25 per cent of the total compensation, which is somewhat smaller than what is required by the reform initiative. Second, the scatter plots of our estimated teacher effects on students' scores against teachers' compensation do not display a positive relationship; there does not appear to have a systematic relationship between teachers' compensation and the estimated teacher effects. Third, our regression analysis shows that after controlling for observed teacher characteristics, course subject, and school dummies, the teacher effects and teachers' compensations are not systematically correlated. In fact, it is interesting to note that the only characteristic that is correlated with teachers' total compensation and base pay is teachers' years of teaching, while performance pay does not appear to be correlated with any teacher characteristics. These results, all taken together, suggest that the performance-based compensation reform is ineffective in terms of incentivising teachers. Given the widespread cross-country evidence that students' reading and math scores increase considerably when teachers are compensated based on students' test scores (Woessmann 2011), a compensation structure that is more closely tied with students' academic performance might be called for.

The paper is organised as follows. Section 2 describes the data and the institutional setting. Section 3 describes our empirical framework and

estimation strategy. Section 4 presents the results of the teacher effect estimation. Section 5 examines the relationship between the teachers' value-addition and teachers' compensation. Section 6 concludes the paper.

2. Institutional setting and the data

2.1 Institutional setting

In China, when students in a typical county enter middle schools, the county enrols students in middle schools near where they live. A general rule can be described as follows. First, the urban area of the county is partitioned into several subareas, each of which has one middle school therein. The middle school enrols all the students who are residents of the subarea inside the urban area. Second, for a rural area, each rural town has its own middle school. Students living in any village of a given rural town are enrolled in the middle school of that town. Third, in most cases, a middle school does not enrol students who are not residents of the associated area (an urban area or a town). Finally, student characteristics such as age, ethnicity and gender are taken into consideration in the assignment of incoming students to different classes to ensure more balanced student composition across different classes. Despite this effort, we find that student composition is not balanced across different classes for certain schools in our data as indicated by the chi-square or Fishers' independence tests. This means that when students are assigned to classes, the school is not able to do so in a completely random manner.

Class organisation of middle school students in the Chinese system differs from that in the U.S. system. In China, students assigned to the same class stay together for all subjects taught and for all grades. This feature makes our estimation of teacher effect immune from the selection bias problem arising from the possibility of students self-selecting into a class/subject. In addition, students tend to be taught by different subject teachers as they move up from grades seven to nine. The three core subjects taught in middle schools in China are Chinese, mathematics and English.

The data set used in this paper is collected from a county in Guangxi province, a southern province of China. While the process of assigning students into different schools and the class organisation manner are the same as a typical county in China, several features make this county an ideal case for estimating teacher effects and consequently examining the incentive role of the teacher performance-based compensation reform. First, the same final examinations (varying by class levels) are used to test students at the end of each semester for all schools in the county. The final examinations are designed and written by the education bureau, the contents of which are not revealed to anybody before the examination time, so none of the middle school teachers and students in the county has any knowledge about the content of the examinations. At the end of each semester, all students are required to take the same final examinations simultaneously. Then, all the

examinations are sealed and conveyed to a specific location designated by the education bureau, where the examinations are graded without students' names. This type of standardised examination process makes the comparison of the test scores of students across schools feasible. Second, because our study concentrates on one county, our estimation of teacher effect is free from the confounding effects of other factors (e.g. curriculum, teacher management and assessment), all of which are similar among all the schools in the same county. Furthermore, the administration officers of the county education bureau demand that teachers from different schools use the same curriculum and that principals from different schools follow the same approach of managing and evaluating teachers.

It is important to note that based on our interviews with school teachers and principals in the study county, we found that subject teachers generally change as students go from grade seven to nine. Unfortunately, we do not have the precise number of cases where a class experienced change of subject teachers throughout the 3-year junior high school education. However, interviews of school principals suggest that approximately 70 per cent of subject teachers change as students go up to the next grade. It is important to point out that our estimation strategy essentially uses the average score of the class after purging the influences of other confounding factors to gauge the teachers' value-addition from a teacher. In fact, this is the strategy of all VAM studies in the literature. Therefore, the identification of teachers' effects does not require subject teachers to be the same across grades of junior high school.

2.2 Data

The data set used in this paper was collected from all the middle school students from one county of a south-west Chinese province for five consecutive semesters (from fall 2008 to fall 2010).[1] The test scores for mathematics, English and Chinese (the core subjects of Chinese middle schools) for all students, and information about students' personal characteristics such as age, ethnic background and sex were collected.[2] In addition, the information on teachers' age, education, gender and their teaching experience was also collected. We then further matched each student in our sample with teachers who are the main instructors of different subjects at

[1] For one particular urban school, we only have data for the students enrolled in fall 2010. Due to the lack of panel data, we excluded this school from our analysis. However, our estimation results are not affected by the exclusion of that particular school. In fact, the student sample size of that particular school (86 students) was only a tiny fraction of the total student sample (10,796) in the county. Specifically, excluding this particular school results in a reduction of 0.8 per cent of the total observations.

[2] It is not possible to have data on test scores for the semester when students graduate from the middle school. This is because the final examination they took during that semester was the high school entrance examination. As a result, for all the students, we have their test scores for up to five semesters: fall 2008, spring 2009, fall 2009, spring 2010 and fall 2010.

Table 1 Student and teacher characteristics in urban and rural schools

	Student characteristics				Teacher characteristics			
	Overall (1)	Urban (2)	Rural (3)	Diff. (4)	Overall (5)	Urban (6)	Rural (7)	Diff. (8)
Age	13.77	13.55	13.87	−0.33*	38.12	39.88	37.04	2.84*
	(1.20)	(1.03)	(1.26)		(7.14)	(8.69)	(5.76)	
Gender	0.55	0.54	0.55	−0.008	0.46	0.37	0.52	−0.14*
(1 = Male)	(0.50)	(0.50)	(0.50)		(0.50)	(0.49)	(0.50)	
Ethnicity	0.52	0.67	0.45	0.22*	0.43	0.60	0.33	0.26*
(1 = Han)	(0.50)	(0.47)	(0.50)		(0.50)	(0.49)	(0.47)	
Class size	41.68	43.94	40.64	3.30*	–	–	–	–
	(7.58)	(6.92)	(7.67)					
Education	–	–	–	–	14.61	14.98	14.39	0.59*
(years)					(1.69)	(1.25)	(1.88)	
Years of	–	–	–	–	16.22	18.17	15.01	3.16*
teaching					(8.00)	(9.49)	(6.66)	
Urban	0.32	1.00	0.00	–	0.36	1.00	0.00	–
No. of	10,712	3,470	7,242	–	214	76	138	–
observations								

Note: ***, ** and * refer to 1%, 5% and 10% statistical significance level, respectively. Standard deviations in parentheses. Numbers may not add as expected due to rounding.

different grades. At the end, the total sample of the data set includes a total of 10,712 students and 27,400 student-semester observations.[3] Since the students' test scores in the data set are tracked and followed over semesters, we are able to estimate teacher effects with the control for influences of the student's prior test scores.

Table 1 presents the summary of the student and teacher characteristics of our data. An average student in the sample is about 14 years old. The sample has 55 per cent of male students and 52 per cent of Han ethnicity. Thirty-two per cent of the students come from urban areas. The class size of an average middle school is 42 students. As for the teachers, they are on average 38 years old. Forty-six per cent of teachers are male and 43 per cent are from majority background. An average teacher has 14.61 years of education and 16.22 years of teaching experience.

The differences in teacher characteristics between the rural and urban schools appear to be pronouncing and tend to support the conjecture that urban schools have better teachers than rural schools. For example, an average teacher in urban schools is almost 3 years older than an average teacher in rural schools, which is also consistent with the fact that the teaching experience of an average urban school teacher is 3.2 years more than that of an average rural school teacher. Similarly, teachers in urban schools

[3] Our sample is composed of several cohorts. For the students who are in grade nine, we have score information for five semesters for each student. For the other cohorts (e.g. in grade seven and eight), we have score data for three semesters for each student. The final sample includes 10,712 students and 27,400 student-semester observations. The sample has no attrition problem.

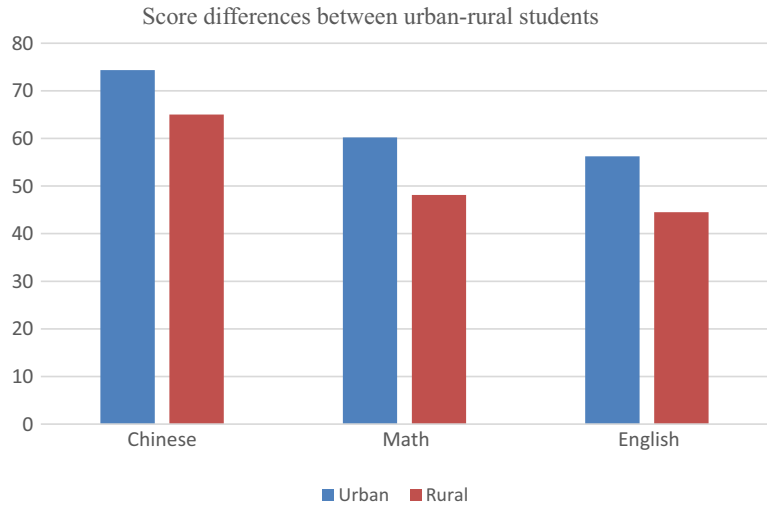

Figure 1 Score differences between urban–rural students. [Colour figure can be viewed at wileyonlinelibrary.com]

are on average significantly more educated than those in rural schools (14.98 vs 14.39). And as in the case of students, the proportion of teachers with the majority ethnicity background is significantly higher in urban schools than in rural schools (60 per cent vs 33 per cent). Teachers in urban schools are also significantly more likely to be female than those in rural schools (63 per cent vs 48 per cent). Except for the gender composition of teachers which has no direct implication for teacher's quality, all the remaining characteristics suggest that urban schools have better teachers than rural schools.

Figure 1 displays score differences between students in urban schools and those in rural schools for Chinese, math and English subjects. Highly consistent with our expectation, there are considerable score gaps between rural and urban schools. More specifically, the scores are much higher for students in urban schools than those in rural schools, and it is consistent across all the three subjects. While not causal, the descriptive evidence tends to suggest that the significant difference in teachers' characteristics may be an important reason for the significant score gaps between the urban and the rural schools. In the next section, we turn to the econometrics analysis to more rigorously estimate the teacher effects on the test scores for each subject and show how the estimated teachers' effects on testing scores are linkd to the incentive-based payment schemes.

3. Empirical model to estimate teacher effects

To empirically estimate the teacher effects on students' academic scores, we employ the standard VAM. The model is commonly used in the literature of education economics to estimate teachers' effects (Aaronson *et al*. 2007; Kane and Staiger 2008; Carrell and West 2010; Sass *et al*. 2014). This approach

essentially attempts to identify the teacher effects on a student's scores, in isolation of other effects on a student's scores related to the individual and family characteristics. In the practice of estimating the value-added, two empirical strategies are employed: (i) randomisation of student assignment into classes and (ii) adding as many relevant and important control variables as possible in the regression to control for the unobservable. In this paper, we follow the second strategy by using observational data to estimate the teacher effects. Others have successfully used observational data to estimate the teacher effects (Jacob and Lefgren 2004; McCaffrey *et al*. 2004; Rockoff 2004; Aaronson *et al*. 2007; Sass *et al*. 2014).

Specifically, we use a straightforward linear dynamic regression model to estimate the teacher effects as below:

$$y_{ijt} = \varphi_g + \omega_t + \beta y_{it-1} + X'_{it}\gamma + P'_{-it}\tau + \mu_j + \theta_i + \varepsilon_{ijt}, \quad (1)$$

where y_{ijt} is student i's test score of a standardised examination. Subscripts j and t stands for teacher j and semester t, respectively. In Equation (1), φ_g and ω_t are, respectively, the grade fixed effect and the time fixed effects. y_{it-1} is student i's standardised score obtained in the previous semester $t-1$. X'_{it} is a vector of student-level variables such as age, ethnicity and gender. X'_{it} are included to control for observable student characteristics. P'_{-it} is a vector of class-level variables that capture the attributes of peer and the classroom effects of student i, including the effect of class size. They are included to control for the texture of the class. θ_i stands for the unobservable student characteristics, such as ability. Finally, ε_{ijt} is the idiosyncratic error term. The key parameter, μ_j, stands for the teacher effect of teacher j, measuring the value-added by teacher j to the students that he/she teaches, with other things being held constant. θ_i, μ_j, ε_{ijt} are assumed to be normally distributed and mutually independent[4]:

$$\mu_j \sim N(0, \delta^2); \theta_i \sim N(0, \rho^2) \text{ and } \varepsilon_{ijt} \sim N(0, \sigma^2).$$

Please note that δ measures the degree of variation or dispersion of the teacher effects.

One common econometrics problem is likely to arise when estimating Equation (1) is the lagged dependent variable (y_{it-1}) is potentially endogenous. Specifically, y_{it-1} is correlated with the error term in Equation (1) as a result of the existence of unobservable student characteristics θ_i. As

[4] This is a standard assumption in the relevant literature. In fact, it is a common assumption in the literature of more general hierarchical models, and the teachers' value-added model is a hierarchical model. The estimation technique for the value-added model was developed under the assumption of normality and independence of these terms. It is not surprising that almost all the teachers' value-added models (Raudenbush and Bryk, 2002; Rockoff, 2004; Jacob and Lefgren, 2008; Rabe-Hesketh and Skrondal, 2012; Carrell and West, 2010) rely on this assumption.

such, the simple ordinary least square (OLS) estimation of Equation (1) leads to biased and inconsistent estimate of β, which could further contaminate the estimation of μ_j, the estimation of which is directly based on the results of β.[5] An instrumental variables (IV) approach is a common approach to obtain consistent estimates in Equation (1) (Sass 2006; Ding and Lehrer 2014).

A standard two-step procedure is used to consistently estimate the teacher effect with an IV approach. First, we employ the generalised method of moments (GMM) with y_{it-2} or longer lags and differences in lagged test scores as instruments to obtain a consistent estimate of β. Second, we purge from the score variable y_{ijt} the variation resulted from the lagged score variable y_{it-1} in the following manner: $y_{ijt} - \widehat{\beta} y_{it-1}$, and we then regress the remained residual of y_{ijt} on the other variables except y_{it-1} to obtain the estimate of teacher effect. Since the remained residual is independent of the influence of y_{it-1}, we are able to control for the endogeneity problem. The detailed description of the estimation procedure is provided in Zhang *et al.* (2018) and Appendix S1.

4. Teacher effect estimation

In this section, we report the estimation of the teacher effects using the VAM. While we use the GMM-IV estimation results as the foundation for the computation of teacher effects, we also report the regression results based on the OLS and fixed effect estimations for comparison reasons. The results are quite revealing as our analysis suggests that the teacher effects account for a significant portion of the score gaps between the rural and the urban students. The results are also robust to different model specifications and assumptions.

4.1 The estimated effect of previous score

Our results indicate that the dynamics of test scores for different subjects are quite different between urban and rural students. For instance, for the English subject, a one-standard deviation increase in the score in the previous semester on average increases a student's current score by 0.691 and 0.469 standard deviations for urban and rural students, respectively. For the Chinese course, the increase is by 0.694 and 0.466 standard deviations for urban and rural students, respectively, although the estimate of rural students is not statistically significant. Unlike the cases of English and Chinese, the results on previous score are not much different between urban and rural students for the subject of Math (0.908 vs 1.132). The detailed regression

[5] It is well known that when the unobserved heterogeneities included in the error term are positively correlated with the explanatory variable, standard OLS estimator without accounting for omitted variable bias is biased upwards (e.g. see Hsiao, 1986). For dynamic panel data, the presence of individual-specific time invariant unobservables could cause the fixed effect estimator to seriously biased downwards in short panels (see Nickell, 1981; Bond, 2002).

Table 2 Estimation of the teachers' value-added model

Variable	Chinese	Mathematics	English
δ	0.222*** (0.020)	0.123*** (0.013)	0.191*** (0.021)
ρ	0.201*** (0.010)	2.78e-12 (8.73e-13)	0.249*** (0.007)
σ	0.569*** (0.005)	0.547*** (0.003)	0.459*** (0.004)
Wald χ^2	865.07	64.44	1086.35
Observation	13,042	12,581	12,410

Note: ***, ** and * refer to 1%, 5% and 10% statistical significance level, respectively. The controlled variables include student age, age squared, gender, ethnicity, class size, class size squared, average age of peer students, average age squared of peer students, male ratio of peer students, ethnicity ratio of peer students, grade fixed effects and year and semester fixed effects.

results and discussion are available in Zhang *et al.* (2018) and the associated Appendix S2.

4.2 The estimated teacher effect

The estimation results of the teachers' VAM for Chinese, math and English show that teachers are a significant contributor to student scores (Table 2). Our results indicate that a one-standard deviation increase in teacher quality (teacher effects) increases student academic achievement by 0.222 (Chinese), 0.123 (math) and 0.191 (English) standard deviations, respectively, for the three subjects. Figure 2 shows the histograms of the estimated teacher effects using the empirical Bayes method. These results suggest that the teacher effect explains a significant part of the variation of student test scores.

4.3 Robustness checks

We also conducted a variety of robustness checks to evaluate the sensitivity of our main results. More specifically, we check whether and the degree to which the results are sensitive: (i) to ethnicity background of the students (Han vs non-Han); (ii) to the effects of after-school tutoring; (iii) to the possible non-random selection problem; and (iv) to other possible residual-error structures of unstructured, autoregressive and moving average. Our robustness checks (not reported) tend to show that the key results are not caused by these observed and unobserved confounding factors. A detailed presentation on the robustness check results and discussion is provided in the robustness check section of Zhang *et al.* (2018).

5. Teacher effects and teachers' compensations

In this section, we examine how closely teachers' compensation scheme is tied to the teachers' value-addition to student academic achievement. We first briefly describe China's teacher performance-based compensation reform, implemented in 2009. With our teacher effect estimates, we then

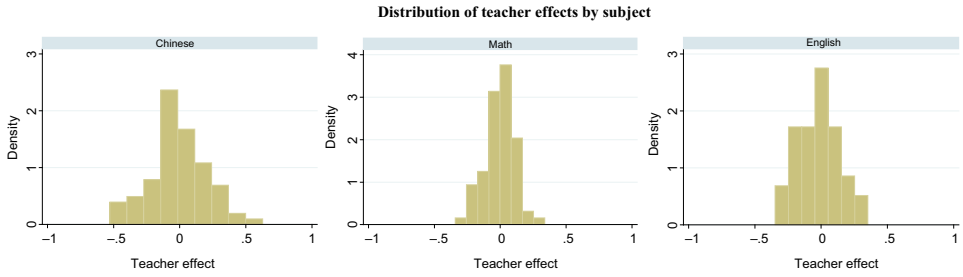

Figure 2 Distribution of teacher effects by subject. [Colour figure can be viewed at wileyonlinelibrary.com]

investigate the relationship between teacher effects and teachers' compensation.

5.1 China's teacher performance-based compensation reform

In 2009, China's Ministry of Education initiated a performance-based compensation reform for elementary and middle school teachers (MoE 2008). The overall goal of this reform initiative was to incentivise teachers for better performance by restructuring teachers' compensation on the basis of their performance. Specifically, teachers' compensation consists of two parts: base pay and performance pay. The base pay accounts for 70 per cent of the total compensation, while the performance pay accounts for the other 30 per cent. The standards of the base pay are developed by the county government or above in accordance with the level of local economic development, price level and teachers' position or rank; these standards are uniform across schools. The performance pay is decided by each school on the basis of each individual teacher's workload and real contribution. In practice, the teacher performance evaluation relies on a variety of quantitative and qualitative indicators, which intend to reflect professional ethics, teacher's workload, effective teaching, relations with colleges and parents, views of colleges and students, participation in the school's research projects and publications, and so on (Lo *et al*. 2013; Wang *et al*. 2014). Several problems, however, have cast doubt on the efficacy of the teacher performance evaluation practice: difficulty measuring and/or quantifying some indicators, the suspected role of school administrators in teacher evaluation due to potential conflicts of interest and the lack of procedural clarity in teacher evaluation.

According to our interviews with the county officials and the middle school principals, the county has closely followed the guidelines of the Ministry of Education and the provincial and municipal-level governments such as dividing the total budget for teachers' compensation into base pay budget and performance pay budget with the former accounting for 70 per cent of the total budget. The standards of the base pay are developed by the county government in a way as described above.

Following these standards, a teacher's base pay is calculated and paid monthly. While the school administration does not have any control over the base salary portion, most of the performance pay budget is at the disposal of the school. Part of the performance pay budget is controlled by the county-level government, which sets up certain subsidy items, for example subsidy items for school principals and rural teachers. The items are implemented uniformly across schools within the county. The school administration sets up a variety of subsidy items, including basic workload subsidy, work quality subsidy, subsidy to support a particular class, administration subsidy, work overload subsidy, education and teaching outcome subsidy. Some items (e.g. basic workload and work overload subsidy) have specific formula to calculate the amount of subsidy paid to each teacher.

Some other items (e.g. the subsidy to support a particular class and the administration subsidy) are position-specific. Other items, such as work quality subsidy, are evaluated by school administrators and peer teachers. Teachers are assessed according to their performance each semester or annually and each teacher receives various subsidy items based on the evaluation results. The performance pay (i.e. various subsidy items) is paid to teachers once at the end of each semester or each year.

5.2 Teacher effects and teachers' compensation

Despite the good intent of the performance-based compensation reform, teachers' compensation is found not to be closely tied to the teachers' value-addition to student scores. First, our data show that after the reform, teachers' monthly compensation increased significantly, almost doubling from 1,589 to 3,108 yuan (Figure 3). The performance pay accounts for 25 per cent of the total compensation, which is somewhat smaller than that required by the reform initiative. Second, the scatter plots of our estimated teacher effects against teachers' compensation do not display a positive relationship (Figures 4–6); there does not appear to be a systematic relationship between teachers' compensation and teacher effects. Third, our regression analyses that further control for observed teacher characteristics, course subject and school dummies show that the teacher effects and teachers' compensation are not systematically correlated (Table 3 row 1). It is interesting to note that the only characteristic that is correlated with teachers' total compensation and base pay is teachers' years of teaching, while performance pay does not appear to be correlated with any teacher characteristics (row 5).

Consistent with our empirical results, our interviews from the field reveal that the budget allocated to each school for performance pay was in large part shared among all teachers. As one of the interviewed school principals put it, 'We have one pie here. As our pie gets bigger as a result

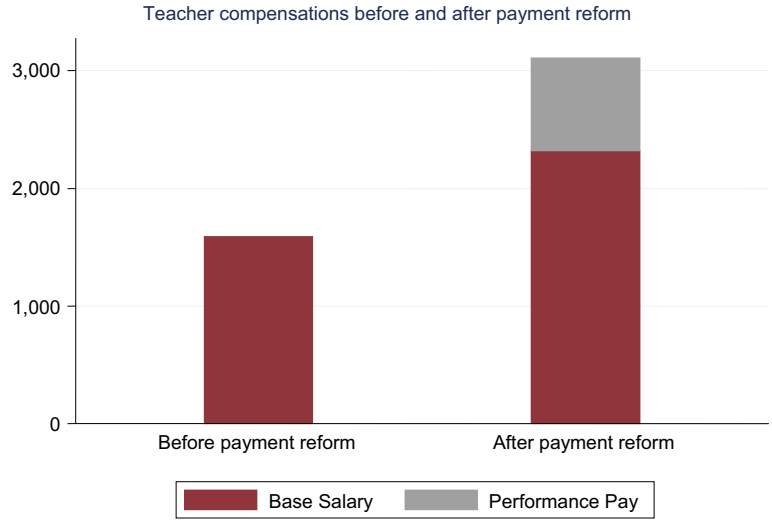

Figure 3 Teacher compensations before and after payment reform. [Colour figure can be viewed at wileyonlinelibrary.com]

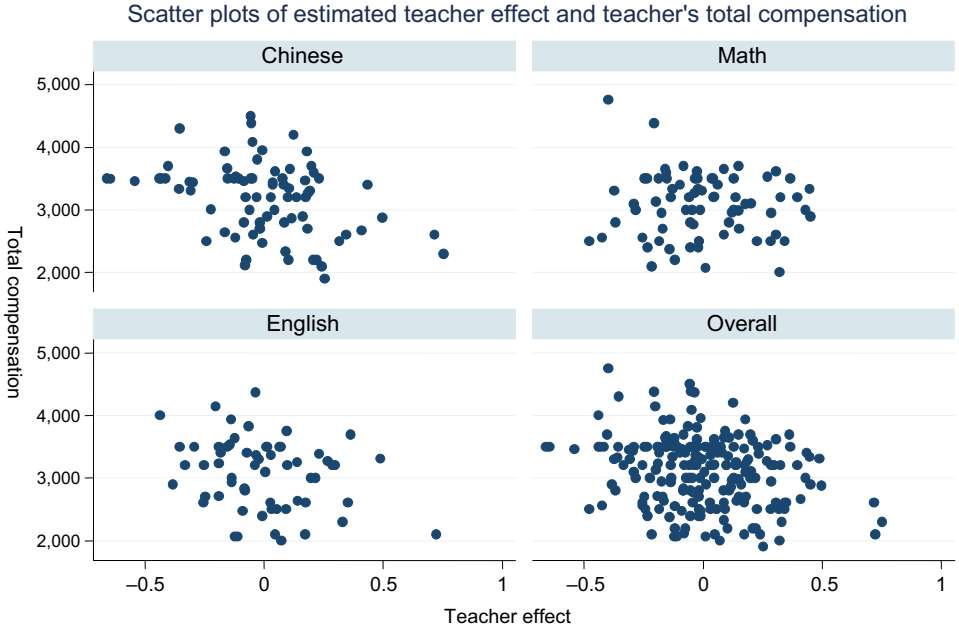

Figure 4 Scatter plots of estimated teacher effect and teacher's total compensation. [Colour figure can be viewed at wileyonlinelibrary.com]

of the reform, we want everyone to benefit from it and share the enlarged pie at the end of each year'. Therefore, the effect of teachers' compensation is likely to be minimal in terms of incentivising teachers and

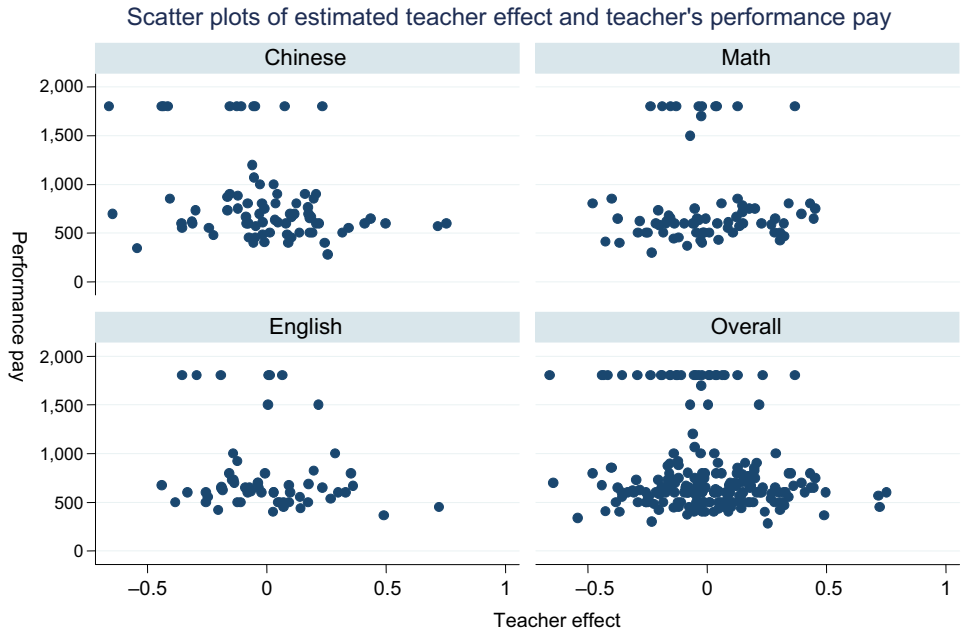

Figure 5 Scatter plots of estimated teacher effect and teacher's performance pay. [Colour figure can be viewed at wileyonlinelibrary.com]

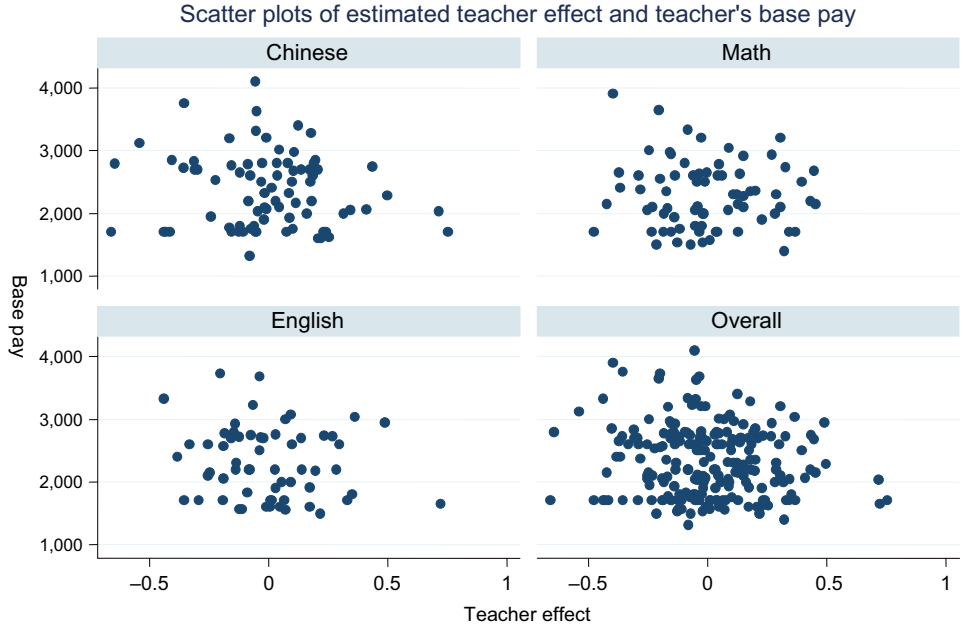

Figure 6 Scatter plots of estimated teacher effect and teacher's base pay. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 3 Correlation of log teacher compensations and teacher effect

	(1) Total compensation	(2) Performance pay	(3) Base pay
Teacher effect	−0.00916 (0.0318)	0.0985 (0.0659)	−0.0392 (0.0453)
Male	0.0270 (0.0294)	0.0603 (0.0408)	0.0156 (0.0313)
Ethnicity (1 = Han)	0.00310 (0.0205)	0.0423 (0.0405)	−0.00422 (0.0193)
Years of education	0.00464 (0.00517)	−0.0107 (0.0113)	0.00848 (0.00479)
Years of teaching	0.0215*** (0.00492)	0.0105 (0.00816)	0.0246*** (0.00537)
Years of teaching squared	−0.000280* (0.000134)	−0.000263 (0.000227)	−0.000293* (0.000151)
Subject dummy	Y	Y	Y
School dummy	Y	Y	Y
Cons.	7.801*** (0.122)	6.520*** (0.225)	7.489*** (0.108)
R^2	0.605	0.784	0.680
Observation	214	214	214

Note: Robust standard errors in parentheses.
* and *** refer to 10% and 10% statistical significance level, respectively. Standard deviations in parentheses.

improving student academic achievement.[6] Given the evidence that across numerous countries, when teacher compensation is based on students' test scores, students' performance in reading and math increase considerably, a compensation structure that ties in more closely with students' academic performance might be called for (Woessmann 2011).

5.3 School principals' opinion and judgement in teacher evaluation

The opinion and judgement of school principals often play a significant role in teacher evaluation. In our survey, we asked school principals to rank teachers' teaching ability in terms of raising student test scores by subject. It is interesting to see how different the subjective evaluation by school principals is from the objective evaluation by our teacher effect estimates. Our results show that the subjective ranking by school principals is only weakly correlated with the ranking by the teacher effect estimates; their correlation is only about 0.4. The scatter plot of the two rankings against each other shows that the data points appear to scatter widely instead of being clustered closely around the 45-degree line (Figure 7). Thus, these results indicate that school principals generally do not clearly recognise teachers' teaching ability and performance and are thus not able to evaluate teachers precisely. As such, the opinions and judgements of school principals in teacher evaluations should be used with caution.

[6] It could be also possible that the teacher compensation reform was implemented in January 2009, and the data we have are up to fall 2010. So it could be too soon for the reform to take effects and be manifested in the data. But at least in the period that our data cover, we do not see a systematic relationship between teachers' compensation and teacher effects.

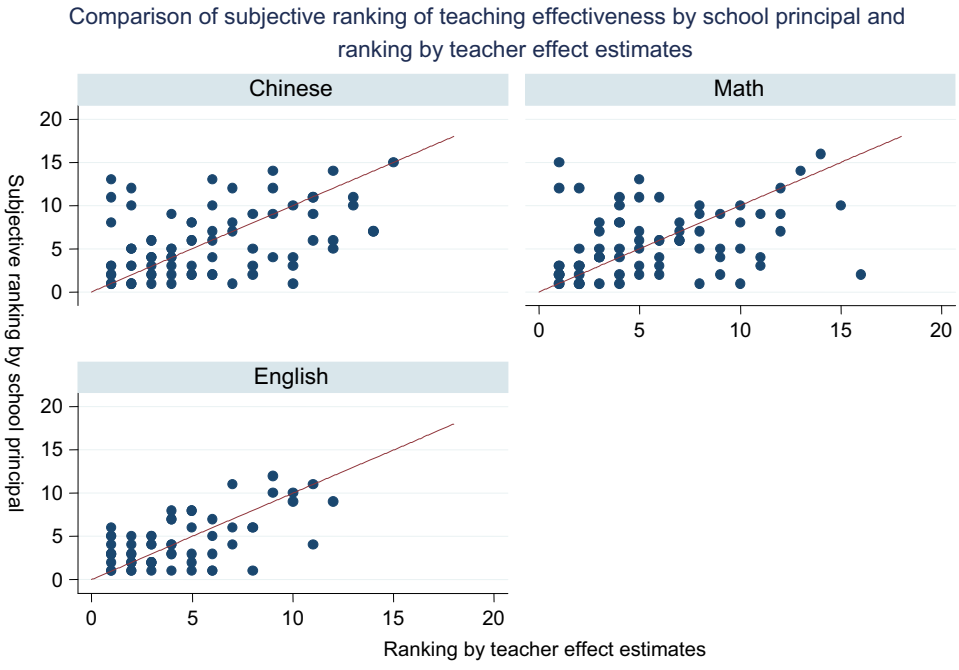

Figure 7 Comparison of subjective ranking of teaching effectiveness by school principal and ranking by teacher effect estimates. [Colour figure can be viewed at wileyonlinelibrary.com]

6. Conclusions

We evaluate the incentive role of the teacher performance-based compensation reform in rural China. Based on a large set of panel data about students and teachers from all the rural and urban schools in one county in Guangxi, we find that the 2009 performance-based compensation reform, which aims to incentivise teachers, is not closely tied to teacher effects. This suggests that the effect of teachers' compensation promoted by the performance-based compensation reform is likely to be minimal in terms of incentivising teachers and improving student academic performance.[7]

There are many factors affecting student academic achievement, such as parenting and school infrastructure and facilities. These factors are hard, if not impossible, to change in a relatively short period. Policies related to teacher compensation, however, might be easier to implement. Better linkage

[7] It is important to note that our results based on a single county in Guangxi are difficult to generalise to the entire country. But on the other hand, focusing on one county gives us a strong internal validity. More specifically, focusing on one county allows us to more precisely estimate the teachers' teaching effects when other confounding factors are more easily controlled. Like many other studies of impact evaluation, we also strike the balance between the internal validity and external validity issue, and the internal validity is the top priority. We think that there is no particular reason to believe the finding from our study would be too different from a more generalised data set because we believe the incentive-based education program is unlikely to be very location specific.

of teachers' compensation to students' academic achievement could greatly motivate teachers and improve student academic achievement. There is a need for further research to identify the best teacher compensation structure that incentivises teachers to perform most.

References

Aaronson, D., Barrow, L. and Sander, W. (2007). Teachers and student achievement in the Chicago public high schools, *Journal of Labor Economics* 25, 95–135.

Barrera-Osorio, F. and Raju, D. (2015). Evaluating the impact of public student subsidies on low-cost private schools in Pakistan, *Journal of Development Studies* 51, 808–825.

Bloom, E., Bhushan, I., Clingingsmith, D., Hong, R., King, E., Kremer, M., Loevinsohn, B. and Schwartz, J.B. (2006). *Contracting for health: evidence from Cambodia*. Brookings Institution, Washington, DC.

Bond, S.R. (2002). Dynamic panel data models: a guide to micro data methods and practice, *Portuguese Economic Journal* 1, 141–162.

Bruns, B., Filmer, D. and Patrinos, H.A. (2011). *Making Schools Work: New Evidence on Accountability Reforms*. The World Bank, Washington, DC.

Carrell, S. and West, J. (2010). Does professor quality matter? Evidence from random assignment of students to professors, *Journal of Political Economy* 118, 409–432.

China Education Statistical Yearbook (ZhongguoJiaoyuTongjiNianjian) (2013). People's Education Press, Beijing.

Ding, W. and Lehrer, S. (2014). Accounting for time-varying unobserved ability heterogeneity within education production functions, *Economics of Education Review* 40, 55–75.

Fryer, R.G. (2013). Teacher incentives and student achievement: evidence from New York City public schools, *Journal of Labor Economics* 31, 373–407.

Gertler, P. and Vermeersch, C. (2012). *Using performance incentives to improve health outcomes, Policy Research Working Paper Series 6100*. The World Bank, Washington DC.

Hanushek, E. and Rivkin, S. (2010). Generalizations about using value-added measures of teacher quality, *American Economic Review: Papers & Proceedings* May, 267–271.

Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge University Press, Cambridge.

Jacob, B. and Lefgren, L. (2004). The impact of teacher training on student achievement: quasi-experimental evidence from school reform efforts in Chicago, *Journal of Human Resources* 39, 50–79.

Jacob, B. and Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education, *Journal of Labor Economics* 26, 101–136.

Kane, T. and Staiger, D. (2008). Estimating teacher impacts on student achievement: an experimental evaluation, NBER Working Paper 14607.

Kane, T., Rockoff, J. and Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York city, *Economics of Education Review* 27, 615–631.

Koedel, C. and Betts, J. (2009). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein Critique, University of Missouri Department of Economics Working Paper 0902.

Lo, L.N., Lai, M. and Wang, L. (2013). The impact of reform policies on teachers' work and professionalism in the Chinese Mainland, *Asia-Pacific Journal of Teacher Education* 42, 239–252.

Loevinsohn, B. and Harding, A. (2005). Buying results? Contracting for health service delivery in developing countries, *The Lancet* 366, 676–681.

Loyalka, P., Sylvia, S., Liu, C.F., Chu, J. and Shi, Y.J. (2019). Pay by design: teacher performance pay design and the distribution of student achievement, *Journal of Labor Economics* 37, 621–662.

McCaffrey, D.F., Lockwood, J.R., Thomas, A.L. and Hamilton, L. (2004). Models for value-added modeling of teacher effects, *Journal of Educational and Behavioral Statistics* 29, 67–101.
MoE (2008). *Guidelines on Performance Evaluation of Teachers in Primary and Secondary schools*. Ministry of Education of the People's Republic of China, Beijing.
Neal, D. (2011). The design of performance pay in education, in Hanushek, E.A., Machin, S.J., and Woessmann, L. (eds), *Handbook of Economics of Education*. North-Holland, Amsterdam, the Netherlands, Oxford, UK, Massachusetts, USA, Vol. 4. pp. 495–548.
Nickell, S. (1981). Biases in dynamic models with fixed effects, *Econometrica* 49, 1,417–1,426.
Nye, B., Konstantopoulos, S. and Hedges, L.V. (2004). How large are teacher effects?, *Educational Evaluation and Policy Analysis* 26, 237–257.
Podgursky, M.J. and Springer, M.G. (2007). Teacher performance pay: a review, *Journal of Policy Analysis and Management* 26, 909–949.
Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata*, 3rd edn. Stata Press, College Station, TX.
Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. SAGE Publications Inc, London, UK.
de Ree, J., Muralidharan, K., Pradhan, M. and Rogers, H. (2015). Double for nothing? The effect of unconditional teachers' salary increases on performance, National Bureau of Economic Research Working Paper No. 21806.
Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005). Teachers, schools, and academic achievement, *Econometrica* 73, 417–458.
Rockoff, J. (2004). The impact of individual teachers on student achievement: evidence from panel data, *American Economic Review* 94, 247–252.
Rothstein, J. (2010). Teacher quality in educational production: tracking, decay, and student achievement, *Quarterly Journal of Economics* 125, 175–214.
Sass, T. (2006). Charter schools and student achievement in Florida, *Education Finance and Policy* 1, 91–122.
Sass, T., Semykina, A. and Harris, D. (2014). Value-added models and the measurement of teacher productivity, *Economics of Education Review* 38, 9–23.
Sylvia, S., Luo, R., Zhang, L., Shi, Y., Medina, A. and Rozelle, S. (2013). Do you get what you pay for with school-based health program? Evidence from a child nutrition experiment in rural China, *Economics of Education Review* 37, 1–12.
Wang, L., Lai, M. and Lo, L.N. (2014). Teacher professionalism under the recent reform of performance pay in Mainland China, *Prospects* 44, 429–443.
Woessmann, L. (2011). Cross-country evidence on teacher performance pay, *Economics of Education Review* 30, 404–418.
Zhang, J., Jin, S., Torero, M. and Li, T. (2018). Teachers and urban-rural gaps in educational outcomes, *American Journal of Agricultural Economics* 100, 1,207–1,223.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Detailed description of the teacher value-added estimation procedure.

Appendix S2. Detailed regression results of dynamics of student test score.