Research Proposal: Estimating Teacher Effect in China

Mo Liu,

supervisor: Professor Atila Abudulkadiroglu

March 2025

1 Introduction

The role of teachers in shaping student outcomes has been extensively studied in the economics of education. Research consistently finds significant variations in teacher effect that cannot be fully explained by observable teacher characteristics such as education level or experience (Kane & Staiger 2008). Instead, teacher effect, measured by their value added to student achievement, captures teachers' unique contribution to educational outcomes. Value-added approach captures the overall impact of a teacher's instruction, classroom management, and other teaching techniques, regardless of whether these traits are formally measured.

Understanding and quantifying teacher effects is crucial for several reasons. First, large-scale studies have shown that students assigned to high value-added teachers not only perform better in standardized tests but also experience long-term benefits, including higher college attendance rates and earnings. Second, quantifying teacher effects can help education systems move beyond credential-based evaluations and toward more performance-based policies.

In the context of China, the topic of the teacher effect is particularly relevant and underexplored. First, the country's highly competitive exam system (e.g., Gaokao) places immense pressure on students and teachers, making teacher quality a key determinant of success, especially at the middle and high school levels. At the same time, there is a significant rural-urban divide in education, with resource disparities leading to concerns about teacher quality in disadvantaged areas. In addition, teacher assignments in China are often determined administratively by school leaders. This makes the Chinese context a policy-relevant setting to estimate teacher effects and understand how teachers impact student outcomes.

This research proposal aims to estimate teacher effects in China using data from the China Education Panel Survey (CEPS), a nationally representative panel dataset that links students to their teachers. This proposal uses a value-added framework to identify how different teachers contribute to student test score gains over time, beginning with a preliminary OLS model and ultimately aiming to implement a fixed effects approach with more data.

The remainder of this proposal is structured as follows. Section 2 reviews the literature on teacher effect, the value-added model and potential sources of biases. Section 3 describes the CEPS data and explains why it is suitable for the proposal. Section 4 outlines the empirical methodology. Section 5 presents preliminary findings based on the available two-year panel. Section 6 concludes and discusses the next steps.

2 Literature review

Over the past two decades, numerous studies have sought to measure teacher quality and estimate its impact on student outcomes. The consensus from these literature is that teacher effect varies significantly and has long-term consequences on students. This section synthesizes key contributions in three main areas: the importance of teacher quality, the estimation of teacher value-added models (VAMs), and the sources of bias in such estimation strategies.

Importance of teacher quality

Foundational work by Rockoff (2004) was among the first to apply panel-data methods to estimate teacher effects, offering early evidence of substantial variation in teacher effects. Building on this foundation, Rivkin et al. (2005) demonstrate that teacher quality explains a large portion of the variation in student test scores,even among students within the same school,suggesting that observable school characteristics alone cannot fully account for differences in student performance. These studies provide compelling evidence for prioritizing teacher quality in education policy.

Estimating teacher value-added

The dominant empirical framework for measuring teacher effects is the value-added model (VAM), which controls student prior achievement and demographics to estimate a teacher's contribution to student growth. Kane et al. (2013) validated this approach using a randomized experiment, showing that well-designed VAMs can reliably capture the causal impact of teachers on student academic outcomes. Furthermore, Chetty et al. (2014) showed that VAMs not only capture student performance in primary schools but also powerfully predict long-term success, such as college attendance and adult earnings, underscoring the broader external validity of the models. In addition, Aaronson et al. (2007) extended the VAM to high school data, affirming their usefulness even in more complex settings.

Source of biases in teacher effect estimation

Despite its strength, the value-added model faces critical methodological challenges. A key concern is the nonrandom teacher assignment to students. Rothstein (2010) demonstrated that value added estimates can be biased, showing that past test scores appear to be influenced by future teachers under VAM estimation. Using North Carolina administrative data, he discovered that student sorting introduces bias even conditional on prior achievement, since classroom assignments are not random. Similarly, Jackson (2014) emphasized that omitting the mechanism behind classroom assignment can yield biased estimates, particularly when schools match students to teachers based on unobserved characteristics. Bacher-Hicks et al. (2014) handled sorting bias using year-to-year changes in teacher assignments within schools, which are plausibly not related to student characteristics. Generally, these studies highlight the importance of accounting for classroom assignment mechanisms to ensure a credible estimation of teacher effects.

In addition to biases from nonrandom sorting, Hanushek & Rivkin (2010) discussed biases arising from measurement error and the challenges of separating teacher effects from classroom peer effects. Although some studies argue that these biases are moderate and manageable (Kane & Staiger 2008), others call for caution in using VAMs for high-stakes teacher evaluations (Rothstein 2010, Jackson 2014, Bacher-Hicks et al. 2014). Ultimately, these papers suggest that while VAMs offer valuable insights, careful model specification and interpretation are necessary.

3 Data

The China Education Panel Survey (CEPS) is a nationally representative longitudinal survey designed to investigate the link between educational outcomes and multiple socioeconomic factors in China. Initiated in 2013, the baseline survey tracks two cohorts of students, 7th and 9th graders, across 112 schools. This study focuses on 7th graders in 2013, a cohort with more exhaustive data and less attrition in subsequent follow-up surveys. The longitudinal structure of CEPS, with planned follow-ups spanning 30 years, offers a unique opportunity to examine the dynamics of teacher effects on student achievement over time.

A critical strength of CEPS for estimating teacher effects lies in its detailed, student-level linkage to specific teachers. For each student, the dataset identifies their assigned teachers in major subjects (Chinese, mathematics, and English), with detailed teacher-level variables capturing teaching experience, level of education, and certifications. Furthermore, CEPS collects standardized midterm exam scores and cognitive ability measures. The former is a low-stakes test as opposed to high-school entrance exam, thus reflecting students' true academic ability. The latter is used to measure students' logical thinking and problem-solving skills, and it is nationally standardized, making it a robust measure of cognitive ability.

The CEPS also provides rich control variables as shown in Table 1, which is important in reducing omitted variable bias and strengthening causal claims about teacher effects. In addition, CEPS could help mitigate endogenous problems like teacher-student sorting, a key challenge in estimating causal teacher effects. Specifically, principals reported school-level tracking practices: 13 of 112 schools explicitly assigned students to classes based on prior academic performance (i.e., tracking), while others used mixed or non-score-based criteria. While non-tracking schools may not achieve perfect randomness, comparing teacher effects across these subsamples allows this study to better understand how sorting biases estimates. Finally, CEPS is a survey conducted based on China's population distribution and other socio-economic factors. The database can be considered representative, making related research findings generalizable to the entire country, providing valuable insights for national education policies.

Table 1: Summary Statistics of 7th graders in 2013

	Full sample		Non-tracking schools		Tracking schools	
	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)
# of students	8984		8125		859	
# of schools	112		99		13	
Chinese teacher	232		189		43	
Math teacher	230		194		36	
English teachers	212		180		32	
Female (%)	47.85		47.83		48.02	
Agricultural hukou (%)	46.37		43.68		71.83	
Only child (%)	44.98		48.23		14.2	
Parent education:						
Below junior high (%)	15.33		14.6		22.08	
Junior high(%)	49.07		48.9		50.65	
Senior high (%)	15.99		15.43		21.13	
Above senior high(%)	19.6		21.04		6.14	
Fiscal appropriation per student	1107.5		1142.6		747.9	
7th grade Chinese score	70.23	9.74	70.14	9.79	71.08	9.20
7th grade Math score	70.21	9.83	70.15	9.85	70.73	9.61
7th grade English score	70.27	9.75	70.17	9.80	71.2	9.22
Cognitive test score	0.028	0.87	0.053	0.87	-0.21	0.78

4 Methodology

This study employs a rigorous econometric framework to identify and quantify the impact of teacher effectiveness on student outcomes, utilizing the classical education production function and addressing critical issues of endogeneity through panel data methods. The methodology is structured in three parts, tailored to the available data structure and the specific context of teacher assignments in China.

4.1 Baseline Specification: Education Production Function

The analysis begins with the classical education production function, which models test scores Y_{iskt} for student i in school s taught by teacher k in year t as:

$$Y_{iskt} = \beta_0 + \mathbf{T}_{kt}\theta + \mathbf{X}_{it}\beta_1 + \mathbf{S}_{st}\beta_2 + \mathbf{F}_i\beta_3 + \epsilon_{iskt} \tag{1}$$

where $\mathbf{T}_{kt}$ denotes teacher-specific inputs (e.g., teacher fixed effect, experience), $\mathbf{X}_{it}$ captures time-varying student characteristics (e.g., prior test scores), $\mathbf{S}_{st}$ represents school-level resources, and $\mathbf{F}_i$ includes fixed student traits. The coefficient θ identifies the average teacher effect under the assumption that $Cov(\mathbf{T}_{kt}, \epsilon_{iskt}) = 0$.

However, this assumption is likely to be violated in practice due to the non-random teacher assignment, a prominent feature in China's education system, where school administrators strategically allocate teachers to classrooms based on unobserved factors (e.g., perceived potential of the student or parental pressure). This administrative discretion introduces correlation between $\mathbf{T}_{kt}$ and ϵ_{iskt}, biasing the OLS estimate of θ.

4.2 Preliminary Analysis: Reduced-Form OLS with Limited Data

Given the current limitation of only two academic years of data (2013-2014 and 2014-2015), a reduced-form OLS regression is first implemented to provide preliminary estimates:

$$Y_{i,8} = \beta_0 + \Sigma_k\theta_k T_{ik} + Y_{i,7}\beta_1 + \mathbf{X}_i\beta_2 + \mathbf{S}_s\beta_3 + \epsilon_i \tag{2}$$

where $Y_{i,8}$ represents student i's test score in his or her 8th grade, $Y_{i,7}$ is defined similarly. $T_{ik} = 1$ if student i is taught by teacher k, 0 otherwise. $\mathbf{X}_i$ and $\mathbf{S}_s$ are student-level and school-level control variables respectively.

This allows for estimating individual teacher effects θ_k, conditional on prior achievement and other control variables. However, for these OLS estimates to be unbiased and consistent, the following conditions must be satisfied:

- $Cov(T_k, \epsilon_i) = 0$. That is, after conditioning on prior test scores and other controls, students are assumed to be quasi randomly assigned to teachers.
- No unobserved sorting: There must be no omitted student or family characteristics that both affect test scores and assignment to certain teachers.

These assumptions are strong and may not hold in practice, especially given the non-random classroom assignment in many Chinese schools. For example, the "tracking schools" reported that they allocate students into different classrooms based on ability, which could create correlation between teacher assignment and unobserved determinants of test scores.

Despite these limitations, the estimated θ_k provide a preliminary measure of teacher effects. To interpret them meaningfully, this study standardizes the test scores within the sample, so that the estimated θ_k represents the effect of teacher k on student test scores measured in standard deviations relative to the average teacher. For example, if $\hat{\theta}_k = 0.15$, this implies that being taught by teacher k is associated with a 0.15 standard deviation increase in 8th-grade test scores, relative to the average teacher, holding prior test scores and control variables constant.

While acknowledging that OLS estimates are likely upward-biased due to non-random teacher-student matching, this preliminary analysis offers a benchmark for comparing subsequent fixed-effects estimates, thereby quantifying the magnitude of selection bias in cross-sectional analysis.

4.3 Addressing Endogeneity: Teacher Fixed Effects Model

A central limitation of the current dataset is that it contains only two years of student-teacher matched data, which restricts empirical strategies to identify teacher effects correctly. To estimate causal effects of teachers while addressing the key concern of endogeneity in teacher assignment, it is essential to obtain a longer panel covering more cohorts or additional school years.

Specifically, with data spanning at least three or more periods, this study would utilize a teacher fixed effects (FE) model, which controls for time-invariant unobserved teacher characteristics and allows for identifying teacher value-added without relying on the strong assumption that teacher assignment is at random. FE approach is widely used in the teacher value-added literature and is considered robust non-experimental strategies to estimate teacher effects while mitigating endogeneity concerns (Rockoff 2004, Chetty et al. 2014).

The fixed effects (FE) model is specified as:

$$Y_{it} = \beta_0 + \tau_k + \alpha_i + \gamma_t + \mathbf{X}_{it}\beta + \epsilon_{it} \tag{3}$$

where τ_k represents teacher k's fixed effect, α_i controls for the time-invariant student fixed effect, γ_t represents the year-specific shock, and $\mathbf{X}_{it}$ is the vector of time-variant control variables (including prior test scores).

A key rationale for introducing teacher fixed effects τ_k is to address potential bias from non-random teacher assignment. In China, classroom assignments are often made by school leaders based on unobserved teacher characteristics, such as reputation and compatibility with certain student groups. If these characteristics are stable over time, including teacher fixed effects controls for such time-invariant unobservables, thus removing selection bias driven by administrative decisions.

In addition, including student fixed effects α_i further strengthens the model by controlling for time-invariant student-level unobservables, such as innate ability, motivation, and family support. These factors are crucial because student-teacher matching may be endogenous not only from the teacher side but also from the student side; higher-ability students may systematically end up with certain types of teachers. Student fixed effects eliminate bias stemming from such sorting mechanisms by comparing changes in outcomes within the same student over time, when taught by different teachers.

5 Results (for preliminary analysis)

This section presents the preliminary analysis of teacher effect on students' test scores across major subjects: Chinese, Math and English (Table 2). The estimation is based on equation (2) that controls for lagged test scores and other control variables.

As expected, the lagged test score ($Y_{i,7}$) is a strong predictor of current performance. The coefficient is large and statistically significant at the 1% level across both subjects, confirming the importance of prior achievement in explaining current outcomes. This result is consistent with the value-added literature and validates the structure of the model.

The results indicate that teacher effect has a substantial impact on student performance. Since test scores are standardized to have a mean of 70 and a standard deviation of 10, the results in Table 2 show that difference in teacher quality translates into an average of approximately 1.5-1.7 points in student test scores across subjects.

To adjust for potential measurement error, this study also report adjusted standard deviations (Adjusted SD) using the maximum likelihood estimation approach. The adjusted estimates are 1.2-1.4 across subjects, suggesting that the true variation in teacher quality remains significant even after accounting for potential attenuation bias.

For comparison with previous studies that normalize test scores to mean=0 and SD=1, we can further standardize these estimates by dividing by 10. This yields teacher effect estimates of 0.130 for Chinese, 0.128 for Math and 0.122 for English, which are consistent with prior literature that typically finds teacher effects in the range of 0.1 to 0.2 standard deviations (Hanushek & Rivkin 2010).

Overall, this preliminary analysis, though limited by endogeneity concerns, suggests the value of a more rigorous teacher fixed effects model that could better isolate causal effects. The observed variation in teacher effect provides strong motivation for obtaining longer panel data and expanding the analysis beyond the current two-year short panel.

Table 2: Preliminary results

VARIABLES	(1) Chinese	(2) Math	(3) English
Teacher Effect(SD=10)			
Raw SD	1.678	1.705	1.510
Adjusted SD	1.304	1.275	1.215
Control variables			
Lagged Chinese	0.422***	0.114***	0.136***
	(0.013)	(0.013)	(0.012)
Lagged Math	0.172***	0.520***	0.144***
	(0.012)	(0.014)	(0.011)
Lagged English	0.241***	0.222***	0.603***
	(0.013)	(0.015)	(0.013)
Other controls	y	y	y
R2	0.6010	0.5976	0.6848
Adj. R2	0.5879	0.5845	0.6752

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

6 Conclusion and next steps

This proposal presents preliminary evidence on the variation in teacher effect in China, using a value-added framework and two year panel data in CEPS. The initial OLS results suggest that teacher assignment is significantly associated with differences in student test scores, even after controlling for prior performance and student characteristics. These findings are consistent with the broader literature highlighting the importance of teacher quality and indicate meaningful heterogeneity in teacher effects.

However, due to the limited time dimension of the current dataset, these estimates remain vulnerable to endogeneity arising from non-random student-teacher assignments. To address this, future work will require access to a longer panel that tracks students and teachers over multiple years. This would allow for the implementation of teacher fixed effects models, which rely on weaker assumptions and can more credibly isolate causal effects.

In the next stages of the research, I plan to improve the empirical design in several directions. First, I will incorporate additional measures of students' performance, such as classroom performance and students' cognitive test scores, which place greater emphasis on reflecting students' thinking abilities, avoiding the drawbacks of content-specific knowledge. Second, I will explore Bayesian shrinkage methods to adjust for estimation noise and small sample bias in teacher effect estimates, following recent methods in the literature.

Moreover, some schools in the CEPS dataset report practices of "tracking" explicitly, which introduces clear non-random assignment mechanisms. With more data, I may leverage regression discontinuity designs (RDD) to better identify teacher effects in settings where assignment rules are clear. For example, if tracking is based on previous test scores, students near the score threshold between top and bottom classes can be considered similar in ability and background, making their classroom assignment approximately random around the threshold. A richer panel would also allow me to study heterogeneous effects, such as how teacher effect varies across school types or student subgroups.

Overall, this research seeks to contribute to the understanding of teacher effect in China and to inform policies that promote more equitable and effective teaching practices. Future extensions, supported by improved data and more robust empirical strategies, will aim to generate insights for education practices and policies.

References

Aaronson, D., Barrow, L. & Sander, W. (2007), 'Teachers and student achievement in the chicago public high schools', *Journal of labor Economics* **25**(1), 95–135.

at Renmin University of China, N. S. R. C. (2025), 'China education panel survey (ceps)'. Accessed: 2025-04-17.
URL: *http://ceps.ruc.edu.cn/English/Home.htm*

Bacher-Hicks, A., Kane, T. J. & Staiger, D. O. (2014), Validating teacher effect estimates using changes in teacher assignments in los angeles, Technical report, National Bureau of Economic Research.

Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014), 'Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood', *American economic review* **104**(9), 2633–2679.

Hanushek, E. A. & Rivkin, S. G. (2010), 'Generalizations about using value-added measures of teacher quality', *American economic review* **100**(2), 267–271.

Jackson, C. K. (2014), 'Teacher quality at the high school level: The importance of accounting for tracks', *Journal of Labor Economics* **32**(4), 645–684.

Kane, T. J., McCaffrey, D. F., Miller, T. & Staiger, D. O. (2013), 'Have we identified effective teachers? validating measures of effective teaching using random assignment', *Research Paper. MET Project. Bill & Melinda Gates Foundation* .

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

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *econometrica* **73**(2), 417–458.

Rockoff, J. E. (2004), 'The impact of individual teachers on student achievement: Evidence from panel data', *American economic review* **94**(2), 247–252.

Rothstein, J. (2010), 'Teacher quality in educational production: Tracking, decay, and student achievement', *The Quarterly Journal of Economics* **125**(1), 175–214.