Journal of Economic Behavior and Organization 214 (2023) 747–781

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Teacher characteristics and student performance: Evidence from random teacher-student assignments in China

Wei Huang[a], Teng Li[b,1,*], Yinghao Pan[c,1], Jinyang Ren[b,1]

[a] National School of Development, Peking University, 5 Yiheyuan Road, Haidian District, Beijing 100871, China
[b] International School of Business Finance, Sun Yat-sen University, Zhuhai, Guangdong 519082, China
[c] National Academy of Development and Strategy, Renmin University of China, 59 Zhongguancun Road, Beijing 100871, China

ARTICLE INFO

JEL Codes:
H75
I21
J24

Keywords:
Teacher quality
Education production function
Student performance

ABSTRACT

This paper investigates the impacts of teacher characteristics on student performance using a randomly assigned teacher-student sample in China. We find that having a more experienced or female homeroom teacher (HRT) with additional classroom management duties significantly improves students' test scores and cognitive and noncognitive abilities. In contrast, these effects are not observed for subject teachers who are responsible only for teaching. More experienced or female HRTs are also associated with a better classroom environment, more self-motivated students, more parental involvement, and higher parental expectations. These mechanisms explain 10–25 percent of HRT effects on test scores and cognitive ability and 50–60 percent of HRT effects on noncognitive ability. Our findings highlight the importance of teacher management skills in education production.

1. Introduction

Since the seminal work by Hanushek (1971), many economists have been interested in identifying and quantifying the effect of teachers on student performance. Although they have reached a consensus that teachers play a vital role in improving education production (Staiger and Rockoff, 2010; Hanushek and Rivkin, 2012; Jackson et al., 2014; Hanushek, 2020), a growing body of literature further explores what and when specific factors may contribute to teacher effects (Jepsen, 2005; Aaronson et al., 2007; Bau and Das, 2020).

Besides the mixed findings from the literature,[1] previous studies mainly focus on teaching skills and largely ignore the role of management of teachers. One exception, Bloom et al. (2015), finds that the leadership of principals accounts for student academic performance and that managerial practices of schools may be an important determinant for the quality of education. In practice, positions in schools may require different dimensions of skills – some may require more management skills and others more teaching skills, or both. Therefore, a couple of natural questions include 1) whether the effects of teacher characteristics vary by position, and 2)

* Corresponding author.
E-mail address: liteng27@mail.sysu.edu.cn (T. Li).
[1] These authors contributed to the manuscript equally.
[1] For example, Krueger (1999) and Rivkin et al. (2005) find that teacher experience has little impact on teacher quality, while both Mueller (2013) and Rockoff (2004) show a significant effect of teacher experience. Antecol et al. (2015) shows that having a female teacher lowers the math test scores of girls in primary school, while Gong et al. (2018) present evidence that having a female teacher has positive and significant effects on girls' performance.

https://doi.org/10.1016/j.jebo.2023.08.024
Received 22 November 2022; Received in revised form 16 June 2023; Accepted 27 August 2023
Available online 8 September 2023

what are the possible mechanisms of the observed effects given different roles of teachers. Answers to these questions can be applied to policy and practice to improve educational outcomes through the processes of recruitment, assignment, and evaluation. For example, understanding how observable characteristics reflect teacher quality in different positions provides administrators with helpful information when hiring.

This paper exploits a unique setting in the Chinese education system to answer the above question. In Chinese junior high schools, each class is typically assigned a homeroom teacher and a few subject teachers. Besides teaching, homeroom teachers are also responsible for class management, including student engagement, communicating with parents, and collaborating with subject teachers. In contrast, subject teachers are solely responsible for teaching tasks on a specific subject. By comparing the effect of the two teacher types on student achievement, we demonstrate the importance of teacher management.

We use a randomly matched student-teacher sample from China and examine the effects of teacher characteristics on student performance, including test scores and cognitive and noncognitive abilities.[2] Our data are from the China Education Panel Survey (CEPS), which contains rich information on teachers (including homeroom teachers and subject teachers), students, and parents from more than 100 junior high schools across mainland China. With ample contemporary information on student, parent, and teacher behaviors and subjective opinions collected by the CEPS, we identify the underlying mechanisms and quantify the contributions of each mechanism.

One concern relates to the validity of random assignment in the *reported* randomly assigned teacher-student sample. The principals and teachers may misreport their method of assigning students. In addition to considering anecdotal evidence,[3] we conduct several statistical tests to alleviate this concern. First, we show that most pre-determined student characteristics, such as parental education, only-child indicators, and previous academic performance, are insignificantly associated with teacher characteristics. Second, in the baseline sample, we randomly assign students to teachers within school-grades and examine the correlations between teacher and student characteristics. The distributions of p-values from the correlation tests in our pseudo-assignments have no significant differences from those in the actual assignments.

We exploit variation from classroom randomization to identify the effects of teacher characteristics, including experience, gender, education level, and college major. Conditional on school-by-grade fixed effects and pre-determined student characteristics, our identification hinges on the variations between classes within each school-by-grade cell, and we examine the characteristics of both homeroom teachers (HRTs) and subject teachers (SBTs). The validity of the randomization in teacher-student assignments ensures consistent estimates in our econometric framework.

Our analysis yields a couple of main findings. First, having an experienced or female HRT significantly increases students' standardized test scores and cognitive and noncognitive ability scores, while having an HRT with an education major or college degree does not significantly affect student performance. Specifically, a one-standard-deviation (SD) increase in HRT experience improves students' test scores by 0.10 SD, cognitive ability scores by 0.12 SD, and noncognitive ability scores by 0.07 SD. Students with a female HRT are expected to have 0.21 SD higher test scores and 0.1 SD higher noncognitive ability scores than those with a male HRT. Our findings are consistent when we examine the results for students in the first year at their school and the next year (i.e., Grade 7 and Grade 8). This suggests that the teacher effects are not short-lived.

Meanwhile, none of the SBT characteristics significantly influences student performance. If teaching skills mattered in the HRT effects, we would expect to observe similar effects for SBTs. However, our results do not support this. Additionally, the impacts of HRT experience are significant in subjects taught not only by HRTs but also by SBTs, with similar magnitudes. If teaching skills mattered in the HRT effects, the impacts would be more salient in the subject HRTs taught. These results imply that the impacts of HRT characteristics may not be purely explained by high teaching skills that have been highlighted in previous literature, such as Chetty et al. (2014). Therefore, we conclude that the role of the position matter when investigating teacher effects.

Finally, we explore the underlying mechanisms and quantify the contributions of different mechanisms to the HRT effects. With the rich information collected by the CEPS, we examine a battery of potential mechanisms, including (1) classroom environment (Lavy and Sand, 2019); (2) student motivation (Heckman et al., 2013); (3) parental involvement, such as parent-teacher interaction (Dizon-Ross, 2019), parental expectations (Cunha et al., 2020), and parental supervision (Malamud and Pop-Eleches, 2011; Gallego et al., 2020); and (4) teacher effort and better match with SBTs (Angrist et al., 2013). Our results show that a one-SD increase in HRT experience improves the classroom environment, student motivation, parent-teacher interaction, parental expectations, and parental supervision by 0.07, 0.08, 0.08, 0.07, and 0.05 SD, respectively. No evidence supports that teacher effort or matching with better SBTs matters. Likewise, students with a female HRT have a better classroom environment, stronger student motivation, and more frequent parent-teacher interaction, all with increases of 0.10 SD, and stronger parental supervision, with an increase of 0.07 SD. In sum, these mechanisms can explain 10–25 percent of the effects of HRT characteristics on test scores and cognitive ability and 50–60 percent of the effects on noncognitive ability. These underlying mechanisms suggest that managerial practices embedded within HRT

[2] The Compulsory Schooling Laws (CSLs) in China, passed in 2006, require that junior high schools not assign students to lower- or upper-track classes based on their pre-entry scores. Although the practice is discouraged, the CSLs allow some schools not to follow randomization depending on local conditions. With permission from their local administration, schools may not be punished for such violations against the CSLs. Based on the information reported by school principals and homeroom teachers, we restrict our main analysis to the randomly assigned teacher-student sample.

[3] When schools claim that they assign students randomly, any violation of the rule would damage schools' reputation. For example, parents could complain to the local education administration and even report the violation to the media, resulting in social pressure on the school. Since teacher bonuses are directly linked to student performance in most cases, teachers have the incentive to include students with better academic records in their own classes. Therefore, it is fair for everyone to follow random assignments. Peer pressure from other teachers may prevent potential changes.

characteristics are an important determinant of student performance.

Our findings contribute to several strands of literature. Above all, this paper contributes to the literature by opening the "black box" of teacher effects on student achievement and builds up the ongoing literature on school management practices. Recent studies suggest that management skills play a crucial role in student achievements (e.g., Kane et al., 2011; Bloom et al., 2015; Tavares, 2015; Lavy and Boiko, 2023). Yet few of them compare the importance of teaching skills with management quality. Comparing the differential effects between HRTs in charge of managing and teaching and SBTs who are only in charge of teaching, we find that the management role of teachers is crucial for teacher quality. However, previous studies usually do not consider the differential impacts of teacher characteristics by teacher responsibilities in schools, thus giving mixed results.[4] By further investigating the mechanisms, including classroom environment and parental involvement, we show that teacher-level management is crucial for the nexus between teacher characteristics and student performance.

Second, our study contributes to the growing literature about teacher effects by investigating their impacts on cognitive and noncognitive abilities besides test scores. Most previous studies focus on its impact on test scores (e.g., Antecol et al., 2015). However, students' cognitive and noncognitive abilities are also essential predictors for life-cycle earnings besides test scores (Kautz et al., 2014). Unfortunately, they receive much less attention in teacher effects literature, especially among developing countries (Evans and Popova, 2016; Jackson, 2018). Although this study is not the first one examining the teacher effects using data from China (e.g., Gong et al., 2018), our research question and conclusions differ broadly from previous literature. Specifically, Gong et al. (2018) focus on how and why female teachers respectively affect the outcomes of boys and girls, thus emphasizing the gender-specific impacts of having a female teacher.[5] Aiming to answer what specific factors may contribute to teacher quality in general, we show the impact of a series of different characteristics of HRTs and SBTs.

The paper proceeds as follows. The subsequent section describes junior high schools in China. Section 3 describes the data. Section 4 tests covariate balance in class assignments. Section 5 presents the results. Section 6 discusses the possible mechanisms for the teacher effects. Section 7 concludes the paper.

2. Background: junior high schools in China

The Compulsory Schooling Laws (CSLs) in China require all age-eligible children (i.e., aged 13–15 years) to complete junior high school education. As the last three years in compulsory schooling, this stage starts in Grade 7 and ends at Grade 9 in most regions. To ensure that every age-eligible child is able to receive an education, governments have established approximately 52,000 junior high schools nationwide as of the end of 2018, i.e., more than 18 junior high schools *per* county. Based on statistics in 2018 from the National Bureau of Statistics of China, more than 99 percent of age-eligible children participate in junior high school education.

As proposed by the CSLs, student enrollment in junior high schools follows the principle of "Division by District and Nearby Admission".[6] Specifically, local administrations divide their administrative regions into several *school districts* and establish specific admission procedures to enroll age-eligible children living in the corresponding school district. For better education quality, parents are likely to move to corresponding school districts by purchasing a local house or apartment (Chan et al., 2020).[7] To rule out the sorting effect, our regressions include the school-grade level fixed effects so that all the effects are identified among the students within the same grade in a school.

Class Assignments. According to the CSLs, junior high schools should not assign students to different ability-based tracks according to their pre-entry scores.[8] To meet the requirements, junior high school administrations generally assign students in one of the following three ways: 1) pure-random assignment, 2) "snake-shaped" assignment, and 3) "first-last" assignment, as illustrated in

[4] Notably, in the context of Chinese senior high schools, Hoekstra et al. (2018) emphasize the importance of "teacher quality" in explaining peer effects. But what is inside the "teacher quality" remains unclear.

[5] Specifically, our paper differs from Gong et al. (2018) in terms of perspectives, underlying mechanisms, and samples. First, Gong et al. (2018) primarily investigate the impact of teacher gender on student outcomes, focusing on gender issues. In contrast, our study explores the roles of HRTs and SBTs in the Chinese education system, emphasizing management practices in junior high schools. Second, Gong et al. (2018) examine mechanisms like teacher questioning and gender stereotypes. In comparison, our study explores classroom environment, student motivation, parental involvement, teacher efforts, and HRT-SBT matching. Finally, we utilize both waves of available data from the CEPS, enabling us to investigate both contemporaneous and persistent teacher effects. By offering distinct contributions, examining different mechanisms, and utilizing a more comprehensive dataset, our study enhances the understanding of how teacher characteristics and teacher management skills impact student outcomes in China.

[6] Article 12 of the CSLs: "The local administration shall ensure that age-eligible children enroll in school nearby their registered residence."

[7] Since school districts are designed on the basis of residence, local administrations allow families without local hukou (i.e., residence registration system in China) to pay a "sponsorship fee" for admission. The amount of this fee depends on the school quality and students' previous performance. For instance, the sponsorship fee in a second-tier city was 7,000 yuan (approximately 1,000 US dollars) in 2014. Chinese news source: http://epaper.cnxz.com.cn/dscb/html/2014-07/09/content_252357.htm.

[8] Before the late 1990s, junior high school quality was evaluated on the basis of student performance on senior high school entrance exams. To increase the senior high school enrollment rate and improve school reputation, junior high schools typically sort students into various tracks by their test scores. However, in the late 1990s, the common practices of sorting students raised concerns about inequality among students. To bridge the growing gap among students of different abilities, the 2006 version of the CSLs emphasizes that junior high schools shall not divide students into an upper track and a lower track based on their previous achievement.

Fig. A1.[9] *Snake-shaped assignment* and *first-last assignment* are also known as balanced assignments. The two methods ensure the student pre-determined characteristics are balanced across classes (Rubin, 1974; Aaronson et al., 2007). For convenience, we call all three assignments *"random" assignments*. The school administration and teachers are incentivized to strictly follow the class assignment results. When schools claim that they assign students randomly, any violation of the rule would damage the school's reputation. For example, parents could complain to the local education administration and even report the violation to the media, resulting in social pressure on the school.[10] Furthermore, teacher promotions and bonuses in these schools rely heavily on student test scores, and therefore, an unbalanced student composition would be unfair for teachers with worse students in competing with other teachers.

However, some junior high schools do assign students to different ability-based classes by tracking.[11] Although this practice is discouraged, the CSLs allow some deviation from random class assignments depending on local conditions. With permission from their local administration, schools may not be punished for such violations against the CSLs. In Appendix Table A1, we regress the random class assignment indicator on the city and school characteristic variables and show that schools in cities with fewer average years of schooling or schools with a lower rank in a city are more likely to use tracking.

Considering the potential matching between teachers and students by tracking, our analysis is restricted to schools that use random assignment methods. Among the 112 schools in our data, 67 percent report that they use such methods.

Class Operation. Class operation in junior high schools follows a homeroom style. A homeroom class is a learning and social unit where most students stay together throughout their three years at the school. Within a homeroom, there are typically 40 to 60 students on average.

Courses and Exams: Students in junior high schools study three core subjects: mathematics, Chinese, and English.[12] For all subjects, students must take at least two standardized exams each semester, including a midterm and a final exam. These exams are organized by school administrations, and all exams are usually marked by the same set of teachers. The test scores are thus comparable across students within one school-grade.

Roles of Subject Teachers and Homeroom Teachers: Each class is allocated a set of SBTs. Typical SBTs have the duties to prepare course materials, give lectures, assign and mark homework and exam papers, and provide course feedback.

For each class, one SBT is appointed as the HRT.[13] When students are assigned to a class, they are typically assigned a homeroom teacher who remains the same for all three years of their middle school education. This educational model, known as "three years, one class," is widely implemented in most junior high schools across China. In addition to taking on the duties of a typical SBT, HRTs are in charge of student engagement, communication with parents, and collaboration with SBTs. Specifically, HRTs maintain homeroom discipline, offer extracurricular activities, and host weekly homeroom meetings. When necessary, HRTs may provide students with one-to-one guidance on matters related to their academic performance, daily behaviors, and emotions. In addition, HRTs exchange information on student performance with parents via phone, social media, and home visits and host face-to-face parent-teacher meetings to provide necessary feedback each semester. Finally, HRTs closely collaborate with SBTs to improve student academic performance.

The roles and responsibilities of HRTs are not exclusive to China, as similar positions exist in other countries as well. For example, in France, an HRT is also known as a *reference teacher (professeur principal)*, responsible for document distribution, advice provision, conflict mediation, and other homeroom tasks. According to Avvisati et al. (2014), in each class, one teacher serves as the "reference teacher," who can be from any subject, but with a higher representation of main subjects like mathematics and French, responsible for being the primary contact for parents and possessing the most in-depth knowledge about the class. Although in other countries (e.g., the U.S.), HRTs may have different roles,[14] school principals, school CEOs, or school counselors take analogous management responsibilities as HRTs in our settings (e.g., Bloom et al., 2015; Mulhern, 2020; Lavy and Boiko, 2023). For instance, Mulhern (2020) suggests that school counselors in the U.S. can have an impact on non-cognitive skills, including behavior and soft skills, by providing mental health counseling, implementing disciplinary measures, and offering general support to help students navigate the challenges of high school.

Appendix Table A2 provides a comparison of teachers with similar responsibilities to HRTs across different countries, shedding light on the global prevalence of these roles. It highlights that the concept of a teacher responsible for guiding and managing a class of students exists in various educational systems worldwide. The information presented showcases the diverse range of countries where teachers fulfill similar roles, emphasizing the importance of understanding and analyzing these positions in a broader international context.

[9] Please refer to Appendix for the detailed descriptions of the three methods.

[10] News source: https://4g.dahe.cn/news/20201015744158.

[11] The tracking procedure can be divided into two steps. In the first step, the school administration sets up a ranking threshold for upper-track students and groups them into one or more classes. If there is more than one upper-track class, random assignment is applied to allocate the top-tier students. Subsequently, the rest of the students are assigned based on random or balanced approaches. By doing so, the schools may attract more promising students for better performance.

[12] Other subjects are political science (Grade 7-Grade 9), history (Grade 7-Grade 9), geography (Grade 7-Grade 8), physics (Grade 8-Grade 9), chemistry (Grade 9), and biology (Grade 7-Grade 8).

[13] A full procedure of class assignments is described in Appendix Figure A1(a). Prior to assigning students to different classes, schools allocate teachers into respective class and assign one of the SBTs as the HRT. .

[14] For instance, HRTs in the US take attendance and make announcements at the beginning of each school day but do not assist students in addressing academic or behavioral issues.

3. Data and summary statistics

Our data are from the two waves (2013–2014 and 2014–2015) of the CEPS.[15] The CEPS, which started in 2013, is a survey of approximately 20,000 students from 112 junior high schools in China. For each sampled school, the first wave surveyed students from *two randomly selected classes,* both in Grades 7 and 9. The second wave collected responses from students in Grade 8 who were also surveyed in Grade 7. Student surveys were distributed and collected in class. In addition to surveying students, the CEPS collected responses from school principals, HRTs, SBTs, and parents. Most data were collected at school, except parental questionnaires were distributed and submitted by students.

As the first step to ensure randomization, we restrict our sample to students at schools that use random assignments based on the information reported by school principals and HRTs. To ensure that the data we use pertains to schools that implement a random classroom assignment policy in grade 7, we rely on survey responses from school principals and HRTs. Specifically, we require that certain conditions are met in our analysis. Firstly, we require the school principal's verification that students are assigned to classrooms in grade 7 using a random approach as introduced in the background. Secondly, the principal must confirm that there is no student rearrangement across classrooms or grades after the initial random assignment in grade 7. Lastly, we also seek confirmation from responses of HRTs teaching in the same grade that students in that grade are not assigned to classrooms based on their test scores or other forms of tracking. Based on our criteria, out of 112 schools, 75 randomly assign students to classes. We exclude students in Grade 9 because the data do not provide class assignment information for their early junior high school stage. These restrictions leave us with 6,742 students in both waves. Table 1 displays the summary statistics for our analysis sample.

Teacher and Class Characteristics. Following standard practice in the literature (Jepsen, 2005; Aaronson et al., 2007), we focus on four observed teacher characteristics: experience, gender, education level, and education college major. Specifically, we define teacher experience as the number of years working as a teacher. We use an indicator for teachers with a bachelor's degree to describe education level and use an indicator for teachers majoring in education to describe teaching qualification. Notably, we use teacher characteristics at initial arrangements (i.e., Grade 7, the year when students were initially enrolled). We do this because some students may have a different HRT in the second year, as the HRTs can leave their position due to diseases, job turnover, or family issues.[16]

Panel A in Table 1 shows the summary statistics for teachers. For HRTs, the average experience in 2013 is 13.7 years; 71 percent are female; 55 percent have a bachelor's degree or above, and 90 percent majored in education. SBTs share similar characteristics with HRTs.[17] Typically, HBTs teach one core subject, and two SBTs are responsible for the other two core subjects for each class. To measure the SBT effects, we take the average of the characteristics of all SBTs in one class in our empirical analysis. A typical class has approximately 46 students in our sample.

Unlike some previous studies focusing on a single characteristic such as gender or experience, we investigate the impact of the four teacher characteristics together. In our sample, the four characteristics are significantly correlated with each other. For example, as shown in Appendix Table A3, female HRTs have fewer years of experience and are more likely to have a college degree. Without controlling all the characteristics in one specification, we cannot isolate the effects of each characteristic.

Measures of Student Performance. We focus on three student performance outcomes: test scores and cognitive and noncognitive ability scores. Detailed descriptions of these measurements are as follows.

Test Scores: Test scores are the most frequently used measure to predict individual future performance. For example, previous literature finds that school test scores positively correlate with human capital and future earnings (see, e.g., Sekhri, 2020). The CEPS collected midterm scores for the three core subjects (i.e., Chinese, mathematics, and English) from school administrations in the fall semester in 2013 and 2014. We standardize the scores for each subject with a mean of 0 and an SD of 10 within each school-grade. We also create another test-score measure of all three subjects by taking the average score, with a mean of 0 and an SD of 10.

Cognitive Ability Scores: General *cognitive abilities,* such as IQ, are crucial in predicting individual earnings (e.g., Heckman et al., 2006).[18] The CEPS conducted a nationwide standardized cognitive ability test assessing students' aptitude in language, visual-spatial abilities, and arithmetic reasoning.[19] The cognitive ability measures are comparable across different schools in each wave,

[15] The CEPS used a multistage probability proportional to size (PPS) sampling design. First, it randomly sampled 28 county and city districts based on their average educational attainment and the percentage of migrants in the population. Next, the CEPS randomly sampled 4 junior high schools in each county or city district. Finally, in each school, the CEPS randomly sampled two classes each from Grades 7 and 9. All students in the sampled classes were surveyed. The response rate was 98.74 percent (CEPS, 2015).

[16] In the robustness analysis, we will show that (1) our results remain unchanged by restricting the sample to students with the same HRTs in Grade 7 and 8 and by using current HRT characteristics, and (2) the teacher effects are persistent in Grade 7 (the first year) and Grade 8 (the following year).

[17] HRTs are chosen by school management based on their characteristics (e.g., management and teaching skills, reputation, and etc). Therefore, HRTs on average may be different from SBTs in some observed and/or unobserved aspects. However, as our identification hinges on the random match between teachers and students, the potential unobserved differences between HBTs and SBTs will not bias our estimates.

[18] Although test scores are commonly perceived as cognitive skills in previous literature, we follow the definition of cognitive skills by Borghans et al. (2016). Specifically, we define test scores as exam outcomes in school (e.g., math scores or reading scores), while defining cognitive abilities as results on general cognitive ability tests such as the Armed Services Vocational Aptitude Battery test (Heckman et al., 2006) and cognitive test scores from surveys (Case and Paxson, 2008).

[19] Specifically, the tests involve questions in (1) language: verbal analogy and verbal reasoning; (2) visual-spatial abilities: visual pattern analysis, origami analysis, and geometry analysis; and (3) arithmetic reasoning: algorithms, sequences, abstract reasoning, probability, and reverse thinking. In practice, the investigators gathered students in Grade 7 (8) to complete 20 (35) questions within 15 (30) minutes.

Table 1
Summary statistics.

	Observations	Mean	Standard deviation
Panel A. Teacher characteristics			
HRTs:			
Experience in 2013 (years)	147	13.71	8.18
Female (=1)	147	0.71	0.45
Bachelor degree or above (=1)	147	0.55	0.50
Major in education (=1)	147	0.90	0.30
SBTs:			
Experience in 2013 (years)	314	15.27	9.61
Female (=1)	314	0.76	0.43
Bachelor degree or above (=1)	314	0.53	0.50
Major in education (=1)	314	0.93	0.26
Panel B. Class characteristic			
Class size	147	45.86	13.21
Panel C. Student performance			
Test score	12,767	0	1
Cognitive ability	12,855	0	1
Non-cognitive ability	11,550	0	1
Panel D. Student characteristics			
Demographics:			
Age in 2013	13,484	12.57	0.72
Female (=1)	13,484	0.48	0.50
Local (=1)	13,484	0.68	0.47
Rural hukou (=1)	13,484	0.50	0.50
Only child (=1)	13,484	0.50	0.50
Han (=1)	13,484	0.90	0.30
Parent bachelor + (=1)	13,484	0.15	0.36
Past experience:			
Rich during childhood (=1)	13,484	0.09	0.28
Kindergarten attendance (=1)	13,484	0.82	0.38
Late enrolment in prim sch (=1)	13,484	0.13	0.33
Transfer in prim sch (=1)	13,484	0.28	0.45
Suspension in prim sch (=1)	13,484	0.05	0.21
Grade skipping in prim sch (=1)	13,484	0.02	0.14
Repeating in prim sch (= 1)	13,484	0.14	0.35

Notes: This table reports the summary statistics of our baseline sample. It includes 6,742 students in 75 schools which randomly assign their students to classes. Academic test scores are the average of three core-subject test scores, standardized with a mean of 0 and an SD of 1. Cognitive and non-cognitive ability are standardized with a mean of 0 and an SD of 1. Past experience is reported by students.

standardized with a mean of 0 and an SD of 10.

Noncognitive Ability Scores: An emerging strand of literature has emphasized the role of *noncognitive abilities*, such as self-control and working hard, on several economic outcomes (e.g., Heckman et al., 2013; Deming, 2017). We follow the approaches by Borghans et al. (2008) and Heckman et al. (2013) to measure noncognitive skills.[20] Additionally, we employ principal component analysis (PCA) to compress the indicators of different aspects into a single index (i.e., the first component), with a mean of 0 and an SD of 10. Appendix Table A4 reports the components and factor loadings of the PCA.

Student Characteristics. The CEPS provides detailed information on student demographics and early-life experiences before junior high school. Specifically, demographic information includes student age, gender, local residence, rural *hukou*, being an only child, Han ethnicity, and having parents with a bachelor's degree. To account for early-life experiences, we use indicators of family wealth during a student's childhood, kindergarten attendance, and indicators of late enrollment, school transfer, suspension, grade skipping, and grade repeating in primary school.[21] Panel D of Table 1 describes student characteristics. In our sample, students' average age in 2013 is 13. Seventy percent of students were born locally, 90 percent are Han ethnicity, and 50 percent are the only child in their families.

[20] Notably, CEPS does not include questions that are directly related to Big Five personality traits. However, we follow Heckman et al. (2013) and consider five components: (1) openness to experience; (2) conscientiousness; (3) extraversion; (4) agreeableness; and (5) neuroticism. According to VandenBos (2007), openness refers to curiosity and intellectual pursuits; conscientiousness means the extent to which people are organized and hardworking; extraversion means the extent to which people are outgoing and sociable; agreeableness means unselfishness and friendliness; and neuroticism relates to the level of depression. In practice, we conceptualize five components: (1) openness to experience: having hobbies, being curious, learning fast, and reacting quickly; (2) conscientiousness: being perceived as hardworking and rarely missing class; (3) extraversion: participating in school activities or frequently going to museums/zoos/science parks or movies/plays/sports with classmates; (4) agreeableness: being friendly and easy-going; (5) neuroticism: indicators whether students have felt depressed, blue, unhappy or meaningless in the past seven days, whether they report self-confidence, and whether they perceive their classmates as friendly.

[21] In Table 1, "Rich during childhood" equals one if students selected "somewhat rich" or "very rich" in their self-reported family wealth before they attended elementary school.

4. Validity of randomization

Our analysis relies on the identification assumption that the sample schools randomly assign their students to teachers. Although we restrict our analysis sample to schools that reported assigning their students to teachers randomly, we cannot rule out the possibility of misreporting. In addition, schools or teachers might manipulate the assignment procedure. Therefore, in addition to the anecdotal evidence described in the background section, we further test whether student assignments are random from a statistical perspective. To this end, we check the validity of randomization in three ways.

Correlations between Teacher and Student Predetermined Characteristics. If the random assignments are appropriately implemented, we would expect that the HRT and SBT characteristics are not associated with student pre-determined characteristics. Similar to Aaronson et al. (2007), we regress the variables for teacher characteristics on each pre-determined student variable while controlling for school-by-grade fixed effects. In columns (1) and (3) of Appendix Table A5, we separately regress each variable of HRT and SBT characteristics on 15 pre-determined student variables and report the p-values of the corresponding estimates. In columns (2) and (4), we regress each of the teacher variables on all pre-determined student variables, using an F-test to show the joint significance of the 15 variables.

With four measures of teacher characteristics (in columns), 15 pre-determined student variables (in rows), and two sets of regressions (separate/joint) for the two types of teachers (HRTs and SBTs) in each panel, we report 240 p-values (4*15*2*2) in Appendix Table A5. Approximately 9 percent (i.e., 22 out of 240) of the coefficients are statistically significant at the 10 percent level. Additionally, none of the regressions are jointly statistically significant at the conventional level. In Appendix Table A6, we restrict our sample to Grade 7 for parallel analysis and find similar results as above. These findings support the reported randomization in student assignments.

Simulation and Comparison. Following Chetty et al. (2009), we implement simulations in the spirit of non-parametric permutation tests. If the actual assignments follow the randomization process, the distribution of p-values in the actual assignments should be close to that in the simulations. Specifically, we randomly assign students to classes within the same grade and school 500 times, keeping the teacher characteristics and class size the same. Next, we repeat the above correlation tests for each pseudo-sample by regressing teacher characteristics on each pre-determined student variable, controlling for school-by-grade fixed effects.

Fig. 1 (a) presents the empirical cumulative distribution function (CDF) of p-values in the correlation tests for both pseudo-samples and our actual sample. In the regime of p-values below 0.1, the CDF in the actual assignments is well below that for the pseudo-samples, revealing that there are fewer statistically significant correlations in the actual assignments than in the simulations. Furthermore, we test the difference between the two CDFs. We calculate the p-value based on a goodness-of-fit test by comparing the two CDFs with the null hypothesis that the two CDFs are identical (Goldman and Kaplan, 2018). We find that the differences in the CDFs between the actual assignments and simulations are statistically insignificant (global p-value = 0.32). When we set the cutoff points of a comparison regime at 0.05, 0.1, or 0.2, the global p-values are 1.00, 0.94, and 0.82, respectively. Reassuringly, the results of the balance tests in the actual assignments are analogous to those in the 500-times simulation, supporting our identification assumption.[22]

Analysis for Schools with Nonrandom Assignment. We repeat the above analysis for schools with nonrandom assignments. If schools manipulate their teacher and student assignments, we would expect more significant correlations between teacher and student pre-determined characteristics. As reported in Appendix Table A7, there are 44 out of 240 coefficients with p-values below 0.1 (i.e., over 15 percent), and five out of eight regressions are jointly significant at the 5 percent level.

Next, using the nonrandom assignment sample, we plot the CDF for the p-values in the correlation tests and compare the CDF to that from random simulations in Fig. 1 (b). The CDF in this sample is significantly higher than that of simulations in the regime of p-values below 0.1. Furthermore, the CDFs between nonrandom assignments and simulations are significantly different across various regimes (global p-value = 0.00). This exercise further validates the reliability of the reported random assignments.

5. Teacher characteristics and student performance

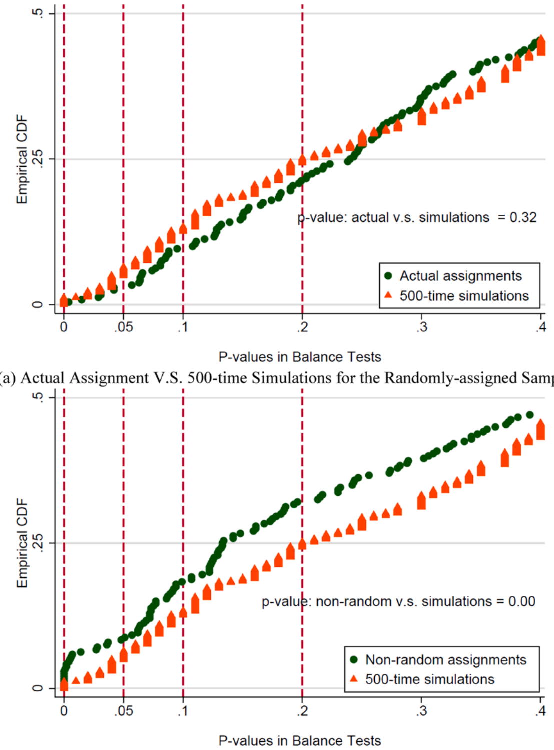
5.1. Econometric framework and baseline results

To quantify teacher effects on student performance, we use the following econometric specification:

$$Y_{icgs} = \beta_0 + \beta_1 HRT_{cgs} + \beta_2 X_{icgs} + \sigma_{gs} + \varepsilon_{icgs}, \quad (1)$$

where Y_{icgs} denotes the outcomes of interest for student i in class c, grade g of school s. The characteristics of HRTs, denoted by HRT_{cgs}, include HRT experience in years, gender (female = 1), education (bachelor's degree or above = 1), and college major (education = 1). The coefficients on these variables, denoted by β_1, capture the impact of HRT characteristics on student performance. Notably, we use

[22] In Appendix Fig. A2, we restrict our sample to Grade 7, plot the CDF of p-values for actual assignments, and compare it with that of the 500-times simulation. The patterns in Fig. A2 show consistent results with the full sample.



(a) Actual Assignment V.S. 500-time Simulations for the Randomly-assigned Sample

(b) Actual Assignment V.S. 500-time Simulations for Nonrandomly-assigned Sample

Fig. 1. Empirical CDF of P-values in balance tests.
Notes: Panels (a)-(b) compare the cumulative distribution of p-values for actual assignment and pseudo-samples. Panels (a)-(b) present the comparison for the randomly- and nonrandomly-assigned sample, respectively. For simulations, we randomly assign students to classes within the same grade and school for 500 times, and repeat balance tests for each pseudo-sample by regressing teacher characteristics on each/all of the pre-determined variables, controlling for school-by-grade fixed effects. The numbers below the charts are global p-values testing the equality of CDF for the actual assignments (panel (a))/ non-random assignments (panel (b)) and 500-time simulations.

initial HRT characteristics (i.e., teacher characteristics in Grade 7 when students were enrolled) to ensure that our results come from the random assignment of HRT to each class rather than other factors. As some students changed their HRTs in the following year, we will show in the later analysis that our results are robust to using current HRT characteristics.

The covariates denoted by X_{icgs} include a set of variables related to student characteristics, including indicators of birth cohort, gender (female = 1), local residence (yes = 1), rural hukou (yes = 1), being the only child (yes = 1), Han ethnicity (yes = 1), and having parents with bachelor's degree (yes = 1) as well as a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60 students). In some specifications, we also include SBT_{cgs}, a vector of variables for SBT characteristics. We control for school-by-grade fixed effects, σ_{gs}, capturing any unobservable factors across school grades. Considering the potential autocorrelation within schools, we cluster the standard errors at the school level.

Table 2 shows the impacts of HRT characteristics on student performance. For each outcome, we report the results following eq. (1). The results in column (1) show that having a senior or female HRT significantly increases student test scores. A one-SD increase in HRT experience (8.18 years) improves student test scores by 0.10 SD (= 0.012*8.18); these results are consistent with those in previous studies on returns to teacher experience (e.g., Harris and Sass, 2011).[23] Next, students with a female HRT are expected to have test scores 21 percent of an SD higher than those with a male HRT. Finally, the effects of HRT educational attainment and college major are statistically insignificant.[24]

Column (2) compares the effects of HRT and SBT characteristics on student test scores. Despite the large magnitudes of coefficients

[23] In Section 5.4, we also consider a non-parametric specification that allows for an added impact due to the nonlinearity of teacher experience.
[24] In a properly executed randomized experiment, the inclusion of student characteristics as the control variables should not change the estimated effects of teacher characteristics in a meaningful way. Appendix Table A8 shows that our estimates are robust without controlling pre-determined student characteristics, suggesting that random assignment was properly executed in our analysis sample.

Table 2
Teacher characteristics and student performance.

	Test score		Cognitive ability		Non-cognitive ability	
	(1)	(2)	(3)	(4)	(5)	(6)
HRT exper.	0.012**	0.011**	0.015***	0.014***	0.008**	0.008**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)
HRT female (= 1)	0.210**	0.206***	0.125	0.121	0.103**	0.107**
	(0.084)	(0.078)	(0.086)	(0.080)	(0.044)	(0.043)
HRT bachelor + (= 1)	-0.058	-0.056	-0.084	-0.082	-0.033	-0.037
	(0.061)	(0.056)	(0.066)	(0.062)	(0.043)	(0.041)
HRT major educ. (= 1)	-0.055	-0.042	-0.098	-0.086	0.051	0.048
	(0.081)	(0.081)	(0.071)	(0.080)	(0.072)	(0.073)
SBT exper.		0.000		0.000		-0.005
		(0.005)		(0.006)		(0.003)
SBT female		0.009		0.010		-0.037
		(0.106)		(0.121)		(0.072)
SBT bachelor +		-0.084		-0.071		-0.016
		(0.109)		(0.116)		(0.084)
SBT major educ.		0.359		0.343		0.064
		(0.245)		(0.270)		(0.156)
p-value for joint test (HRT)	0.09	0.06	0.01	0.01	0.04	0.03
p-value for joint test (SBT)		0.50		0.65		0.59
Observations	12,767	12,767	12,855	12,855	11,550	11,550
R-squared	0.28	0.28	0.34	0.34	0.18	0.18
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of teacher characteristics on student test scores and cognitive and non-cognitive ability scores. The dependent variables are test scores (columns 1 and 2), cognitive ability score (columns 3and4), and non-cognitive ability score (columns 5 and 6). In each column, we also report p-values for joint tests of four HRT or SBT characteristics. Academic test scores are the average of three core-subject test scores, standardized with a mean of 0 and an SD of 1. Cognitive and non-cognitive ability are standardized with a mean of 0 and an SD of 1. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes = 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

on SBT major, none of the coefficients on SBT characteristics are statistically significant at the conventional level, indicating SBT characteristics have limited predictive power for student test scores. Additionally, the coefficients of HRT characteristics are virtually identical to those in column (1).

Columns (3) and (5) show that having a senior or female HRT improves student cognitive and noncognitive abilities. Specifically, an increase of 1 SD in HRT experience raises cognitive and noncognitive ability scores by 0.12 SD and 0.07 SD, respectively. Students with a female HRT are expected to perform 0.1 SD better in noncognitive abilities than those with male HRT. The inclusion of SBT characteristics in columns (4) and (6) does not materially change the estimated impacts of HRT characteristics. As in columns (1) and (2), none of the SBT characteristics significantly impacts student cognitive or noncognitive abilities.

5.2. The impacts of teacher characteristics by subject

Table 3 summarizes the effect of HRT characteristics on test scores by subject. As indicated in columns (1)–(3), the experience of the HRT has a significant impact on Chinese and math test scores, but the effect is statistically insignificant for English test scores. Specifically, a one-SD increase in HRT experience improves Chinese and math test scores by 0.08 SD (= 0.010*8.18) and 0.11 SD (= 0.013*8.18), respectively. In other words, the impact of HRT experience on student outcomes varies across subjects, with a greater influence observed on calculation abilities compared to language skills. This finding is in line with previous studies such as Hanushek and Rivkin (2010), Wiswall (2013), and Papay and Kraft (2015) that show that teacher experience affects math scores more than reading scores. Additionally, our results demonstrate that being assigned to female HRTs influences student performance across all subjects.

Next, we investigate whether HRT experience matters only in the subject that HRTs teach. One potential explanation for the salient HRT effects and muted SBT effects is that HRTs' teaching skills are better than SBTs'. In the analysis, we interact HRT experience with an indicator of an HRT's subject. If the HRT's subject drives the effect of HRT experience, we would expect the interactions to be

Table 3
HRT characteristics and student test scores, by core subject.

	Chinese		Math		English	
	(1)	(2)	(3)	(4)	(5)	(6)
HRT exper.	0.010*	0.012**	0.013**	0.013**	0.008	0.006
	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
HRT female (= 1)	0.146*	0.138	0.194**	0.229***	0.222***	0.178***
	(0.083)	(0.083)	(0.077)	(0.075)	(0.077)	(0.065)
HRT subject (= 1)		0.124		0.001		0.169
		(0.127)		(0.209)		(0.128)
HRT exper. *		-0.006		0.008		0.000
HRT subject (= 1)		(0.008)		(0.011)		(0.008)
Observations	12,790	12,790	12,796	12,796	12,778	12,778
R-squared	0.29	0.29	0.18	0.19	0.29	0.30
Other HRT characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of HRT characteristics on test scores across core subjects. The dependent variables are standardized test scores of Chinese, Math, and English, with a mean of 0 and an SD of 1. Columns (2), (4), and (6) include the interaction between the indicators of the subject taught by HRTs and HRT experience. Control variables include HRT education level and college major, student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30–49, 50–59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

positive and significantly different from zero. However, as the estimates in columns (2), (4), and (6) show, the interactions are not statistically significant for Chinese, math, or English scores. For Chinese scores, the sign of the interaction is even negative. The results indicate that HRT teaching skills may not explain the HRT effects.[25] One possibility is that experienced HRTs may teach students in terms of learning methods via their daily management practices, which may promote students' scores in all subjects.[26]

5.3. Heterogeneous impact of teacher characteristics

Grade 7 / Grade 8. Panels (a) and (b) in Fig. 2 show the effects of HRT experience and gender on student performance by grade, respectively. The effects are similar and do not present a statistically significant difference in the two subsamples. These results suggest that the HRT effects are persistent. Appendix Table A10 shows the corresponding OLS estimates for Fig. 2. Additionally, considering that students are not exactly the same in the two grades, we further restrict our analysis to a balanced-panel student sample, and we observe no material changes. We show the results in Appendix Fig. A3.

Student Gender. Panels (c) and (d) in Fig. 2 investigate the heterogeneous effects by student gender. Although the coefficients are slightly larger for male students, comparisons between the two subgroups reveal statistically insignificant differences at the conventional level.

Rural / Urban. Panels (e) and (f) in Fig. 2 demonstrate that the HRT effects vary by school location (urban or rural areas). We find large and statistically significant impacts of having senior or female HRTs in urban schools and small and insignificant effects in rural schools. These findings are in line with the literature demonstrating that teacher effectiveness in poor regions is lower than that in rich regions (Sass et al., 2012).

5.4. Other results

Results Based on Nonrandom-assigned Sample. It would be helpful to estimate the impacts of teacher characteristics on student performance by using only schools with nonrandom assignments. Such a practice allows a comparison with those from the randomized sample and clarifies the extent of bias in the estimated effects of teacher characteristics from non-experimental data. Appendix Table A11 presents the effects of HRT characteristics among schools in which students and teachers are nonrandomly matched. Compared with the estimates in Table 2, the HRT effects in the nonrandom-assigned sample are even larger. Moreover, the

[25] Similar to the specifications in columns (2) (4) (6) of Table 2, we also include SBT variables in Appendix Table A9, and the basic results are robust. It is noteworthy that the impacts of some SBT characteristics (the proportion of SBTs majoring in education) are significant on subject scores. Therefore, we cannot fully rule out the impacts of SBT characteristics on student performance. However, this does not affect the interpretation of the main results.

[26] Notably, we are not able to disentangle the effects of teaching experience (includes the experience of a teacher serving as SBT) of a current HRT from the effects of the HRT experience. We can imagine that an HRT with more classroom management experience certainly has more teaching experience. Therefore, the results here are suggestive because we could not distinguish whether the effect comes from the interacted effects of past experience and the current status of an HRT or from pure management experience as an HRT. As a further check, we also investigate whether HRT experience matters only when HRTs teach math, Chinese and English and confirm that this is not the case.

Fig. 2. Heterogeneous impacts of HRT characteristics.
Notes: Panels (a)–(f) present the heterogeneous effects of HRT characteristics based on equation (1), by student grade (a-b), student gender (c-d), and type of location (e-f). In each Figure, Panels A–C demonstrates test score, cognitive ability, and noncognitive ability, respectively. Control variables include HRT education level and college major, student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects.

effects of HRTs with a bachelor's degree become statistically and economically significant. These results suggest that students and teachers are positive matchings in the nonrandom-assigned sample. In other words, if we include nonrandom-assigned classes in our analysis, our estimates would be biased upward. We also compare the estimates based on pure-random and balanced assignments and find no significant difference (see Appendix Table A12).

Robustness to Additional Controls and Different Settings. Appendix Table A13 shows the robust results when we control for student early-life experiences. In addition, Appendix Table A14 shows consistent results when we use an alternative approach to construct a noncognitive index suggested by Kling et al. (2007) and when we decompose noncognitive abilities into different components.

Appendix Table A15 shows that our results remain stable if we reweight the test scores across subjects.

Attrition in HRT-student Pairs. It is infrequent that schools make class adjustments. But it is important to consider the changes in HRT-student pairs after Grade 7, as this could raise concerns about potential teacher-student sorting. Specifically, the composition of classes formed through random assignment in Grade 7 might change in Grade 8 through various means. Out of the total 6,742 students in the study, 8.5 percent were not followed up with, and only 0.06 percent underwent switching within schools. Student attrition encompassed school changes for 6.3 percent of students, dropout for 1.2 percent, and other reasons for 0.9 percent. Class reorganization may occur for various reasons, such as changes in school leadership or educational policy. Approximately 6.9 percent of students experienced class reorganization, resulting in 7.6 percent having a different HRT in Grade 8.

Appendix Table A16 presents the relationships between these attrition reasons and HRT characteristics. No systematic change appears in the relationship between HRT characteristics in Grade 7 and attrition in HRT-student pairs in Grade 8. Considering that 7.6 percent of students had a different HRT, we further test the robustness of the results in Appendix Table A17. In Panel A, we drop the students with a different HRT in the following year and find consistent results. In Panel B, rather than using the characteristics of initial teachers (i.e., Grade 7), we use those of current teachers, and our results are robust.[27]

Nonlinearity. The relation between teacher experience and student performance (especially test scores) is not necessarily linear (e.g., Rockoff, 2004). Fig. 3 presents binned scatter plots of residualized teacher experience against student performance for HRTs. As shown, the impacts do not fade out beyond the early stages of teaching careers, in accordance with Harris and Sass (2011) and Gerritsen et al. (2017). As a comparison, the impacts for SBT experience are statistically insignificant, with much flatter fitted linear lines, as shown in Appendix Fig. A4. Besides, Appendix Fig. A5 presents the non-parametric estimates for the effects of HRT experience using a set of categorical variables. Specifically, we divide HRT experience into seven categories, every five years per category. Again, the effects are still salient beyond the early stages of teacher careers.

Analysis of Variance. An important question raised by previous literature is to what extent observed teacher characteristics can explain teacher effectiveness; the magnitude of these effects can indicate whether recruitment and compensation should be based on these characteristics (Rockoff, 2004). We follow previous literature (e.g., Aaronson et al., 2007) and estimate the proportion of the variance of teacher effectiveness that the observed characteristics can explain. Specifically, we first derive HRT effects by regressing each student performance variable on HRT indicators. Then, the analysis of variance (ANOVA) is applied to calculate the explanatory power by dividing the explained variance of each teacher characteristic by total variance. As reported in Appendix Table A18, the total explanatory power of teacher characteristics is mainly driven by HRT characteristics. For example, HRT characteristics explain 7.4 percent of the variation in test scores, which is larger than the findings of 5 percent in Pakistan (Bau and Das, 2020) and only 1 percent in the US (Aaronson et al., 2007). In the next section, we further investigate why HRT characteristics matter. However, one caveat of this analysis is that the variance of HRT effects may not be a consistent estimate, as they are all estimated with noises, including peer effects and other shocks to classrooms.

6. Mechanisms

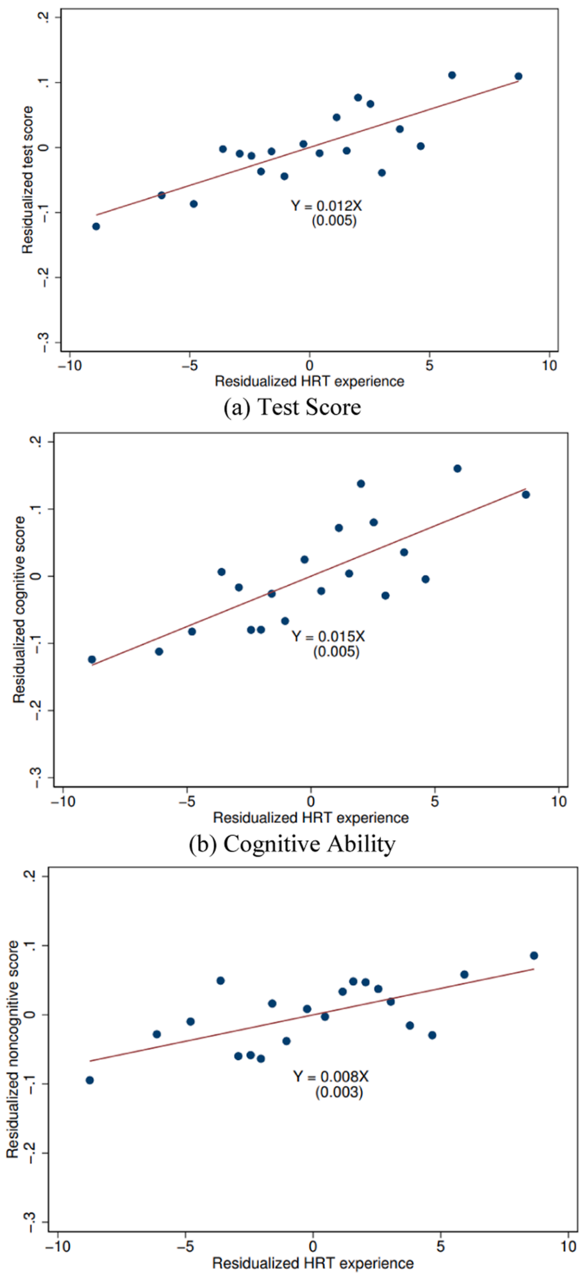
6.1. Potential mechanisms

Given that teacher effects are not significantly explained by HRT teaching skills, this section explores potential mechanisms through which HRT characteristics affect student performance. Specifically, we focus on four types of mechanisms:

(1) *Classroom environment*. HRTs can affect the classroom environment, such as increasing peer quality and peer interaction, through classroom management and activity provision. In a better classroom environment, student performance can improve (Lavy et al., 2012; Lavy and Sand, 2019).
(2) *Student academic motivation.* Academic motivation, apart from its obvious link to performance in school, has been shown to be a powerful predictor of decreased student misbehavior (Heckman et al., 2013). As HRTs may enhance student motivation through daily management practices, they may consequently improve student academic performance.
(3) *Parental involvement.* HRTs regularly share information and communicate with parents about their children's performance and guide parents to support their children in their schoolwork. Thus, HRTs can affect parental involvement through teacher-parent interaction, parental expectations, and supervision of students (e.g., Cameron and Heckman, 2001; Gallego et al., 2020).[28]

[27] Due to the small number of homeroom teacher changes, however, the data do not provide enough statistical power to verify the effects of switching teachers in the sample. Therefore, we explore the potential impact of changes in HRT-student pairs by analyzing the effects of current teacher characteristics on student outcomes. This approach allows us to assess whether changes in the HRT-student pair significantly hange our baseline results.

[28] Frequent parent-teacher interaction, via information provision, improves student academic achievement (Dizon-Ross, 2019; Islam, 2019) and behavior (Avvisati et al., 2014; Gallego et al., 2020). By interacting more with HRTs, parents may raise their expectations of their students. An ongoing strand of literature has pointed out that parental expectations of returns to schooling are critical in student development (Cameron and Heckman, 2001; Cunha et al., 2020). In addition, if parents are well connected with HRTs, they are expected to effectively supervise their children at home. In some daily activities, such as computer usage, parental supervision plays a crucial role in mitigating the negative effects on student achievement (Malamud and Pop-Eleches, 2011; Gallego et al., 2020). In sum, HRTs can elicit more parental involvement via parent-teacher interaction, parental expectations and supervision and subsequently improve student performance.



(a) Test Score

(b) Cognitive Ability

(c) Noncognitive Ability

Fig. 3. Impacts of HRT experience on student performance.
Notes: Panels (a)-(c) present the nonparametric relation between HRT experience and student performance. We residualize teacher experience and each student outcome with respect to HRT gender, education level, major, class size, student demographic characteristics, as well as school by grade fixed effects. Each scatter denotes the mean of the student performance against the mean of residualized teacher experience within each of 20 bins. The solid line is for a linear regression fit on the plotted scatters.

(4) *Teacher efforts and HRT-SBT matching.* As shown by Angrist et al. (2013) and Lavy (2020), increased instructional time can effectively improve student performance. Additionally, senior or female HRTs may be matched to better SBT teams, possibly improving student performance (e.g., Andersen et al., 2020).

We use rich information from the CEPS to measure the outcomes mentioned above. Each single-dimensional mechanism index is consolidated by the first PCA component, with a mean of 0 and an SD of 1. Appendix Table A19 presents the measurements of each mechanism and the PCA results. Appendix Table A20 summarizes the descriptive statistics for each mechanism component.

6.2. Methodology and results

Methodology. To quantitatively estimate the contributions of different mechanisms to the HRT effects, we conduct the analysis in two steps.[29] First, for each mechanism j, we replicate eq. (1) by replacing Y_{icgs} with the potential mechanism index, m^j:

$$m^j_{icgs} = \alpha^j_0 + \alpha^j_1 HRT_{cgs} + \alpha^j_2 X_{icgs} + \sigma^j_{gs} + \varepsilon^j_{icgs}. \tag{2}$$

Second, following Heckman et al. (2013), we include the mechanism indices m^j_{icgs} in eq. (1):

$$Y_{icgs} = \gamma_0 + \gamma_1 HRT_{cgs} + \sum_j \varphi^j m^j_{icgs} + \gamma_2 X_{icgs} + \sigma_{gs} + \varepsilon_{icgs}. \tag{3}$$

Given a specific HRT characteristic r, Gelbach (2016) proves that $\widehat{\beta}_{1,r} = \widehat{\gamma}_{1,r} + \sum_j \widehat{\varphi}^j \widehat{\alpha}^j_{1,r}$. In particular, r takes a value of *experience* or *female* to denote the corresponding coefficient for HRT experience or gender (female = 1). The term, $\widehat{\varphi}^j \widehat{\alpha}^j_{1,r}$, implies the effects of HRT characteristic r on outcomes through mechanism j. Therefore, $\widehat{\varphi}^j \widehat{\alpha}^j_{1,r} / \widehat{\beta}_{1,r}$ can be viewed as the proportion of effects of HRT characteristic r attributable to mechanism j. However, it should be noted that the the variables, m^j_{icgs}, are endogenous and thus the calculation in this section just provides suggestive evidence.

Mechanism Decomposition. We first follow eq. (2) to estimate the impact of HRT characteristics on potential mechanism variables.[30] The results in column (1) of Table 4 imply that a one-SD increase in experience improves the classroom environment by 0.07 SD (= 0.009*8.18). Similarly, the results in other columns suggest that a one-SD increase in HRT experience corresponds to increases of 0.08 SD in student motivation and parent-teacher interaction, 0.07 SD in parental expectation, and 0.05 SD in parental supervision. Likewise, students with female HRTs are predicted to have a better classroom environment, higher motivation, more parent-teacher interaction, and stricter parental supervision than those with male HRTs. Consistent with our baseline findings, there is no significant evidence that the coefficients of HRT education level or college major are statistically significant at the conventional level. Collectively, these results show that a senior or female HRT boosts the classroom atmosphere, enhances student motivation, and gets parents more involved. Appendix Table A21 reports consistent results using an alternative method by Kling et al. (2007).

In columns (1) and (2) of Table 5, we examine the potential impacts of HRT characteristics on teacher effort, measured by working hours. The results are statistically insignificant at the conventional level. Next, we examine whether a senior or female HRT is likely to have a team of SBTs with specific characteristics in columns (3)–(6). The results do not support this.[31]

We further quantify HRT effects by separately calculating the explanatory power of different mechanisms.[32] Panel (a) in Fig. 4 presents the estimated decomposition of HRT experience effects into a wide variety of mechanisms, including classroom environment, student motivation, parent-teacher interaction, and parental supervision. "Unexplained" refers to the residual effect associated with unobservable mechanisms.

Among the three outcomes for student performance we examine—test scores, cognitive abilities, and noncognitive abilities—HRT experience accounts for 23.4 percent, 12.9 percent, and 60.9 percent of the variance, respectively. Overall, experienced HRTs affect student academic performance and cognitive ability mainly through the classroom environment channel. Panel (b) in Fig. 4 shows the results for the effects of female HRTs. Female HRTs mainly affect student academic performance by enhancing academic motivation (6.9 percent), but they improve noncognitive abilities via more intensive parent-teacher interaction (17.3 percent).

7. Conclusion

This paper investigates the role of teacher management skills in the education production function. Using a randomly assigned teacher-student sample in China, we show that senior and female HRTs effectively improve students' test scores, cognitive skills, and noncognitive skills. Specifically, students who have an HRT with one SD of additional experience are predicted to have test scores that are 0.10 SD higher, cognitive ability scores that are 0.12 SD higher, and noncognitive ability scores that are 0.07 SD higher. Meanwhile, students with a female HRT are expected to have 0.21 SD higher test scores and 0.1 SD higher noncognitive ability scores than those with a male HRT. In contrast, SBT characteristics have no impact on student performance.

We further examine the mechanisms of classroom atmosphere, student motivation, parental involvement, and teacher effort/HRT-SBT matching. The identified mechanisms explain approximately 23.4 percent, 12.9 percent, and 60.9 percent of HRT experience effects on student performance, respectively. Likewise, the mechanisms explain approximately 15.0 percent, 16.9 percent, and 49.9 percent of HRT gender effects on student test scores and cognitive and noncognitive abilities.

[29] Notably, the identification assumption here is the same as the baseline analysis, i.e., the random match between students and teachers.

[30] "Mechanisms" here could also be directly interpreted as the effects on classroom environment, student academic motivation, parental involvement, effort, and HRT/SBT matching, as a complementary to the test scores/cognitive/non-cognitive skills in the baseline results.

[31] For consistency, the analysis is conducted at the student level. Indeed, the student-level results in Table 5 are equivalent to teacher-level results weighted by the number of students an HRT manages. In Appendix Table A22, the analysis departs from the student level to be at the HRT level, and we still could not find any significant results.

[32] Following the methodology mentioned above, we estimate equation (3) and report the results in Appendix Table A23. For statistically insignificant mechanisms in Tables 4 or 5 (i.e., coefficient insignificant at the 10 percent level), we do not include them in the equation.

Table 4
Impacts of HRT characteristics on mechanism indices.

	Classroom environment (1)	Student motivation (2)	Parent-teacher interaction (3)	Parental expectation (4)	Parental supervision (5)
HRT exper.	0.009**	0.006**	0.010*	0.009***	0.006***
	(0.004)	(0.003)	(0.006)	(0.003)	(0.002)
HRT female (= 1)	0.104**	0.089**	0.186***	0.064	0.069**
	(0.046)	(0.041)	(0.064)	(0.043)	(0.030)
HRT bachelor + (= 1)	-0.046	-0.043	-0.157	-0.025	0.043
	(0.050)	(0.045)	(0.103)	(0.051)	(0.048)
HRT major educ. (= 1)	-0.093	0.026	-0.199	0.037	0.050
	(0.086)	(0.054)	(0.127)	(0.052)	(0.043)
Observations	12,592	12,232	12,286	12,518	12,733
R-squared	0.13	0.16	0.22	0.13	0.08
Student controls	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the impacts of HRT characteristics on mechanism indices. In columns (1)–(5), the dependent variables are indices of (1) classroom environment, (2) student motivation, (3) parent-teacher interaction, (4) parental expectation, and (5) parental supervision. The dependent variables are single-dimensional measures using PCA to consolidate the mechanism components, normalized with a mean of 0 and an SD of 1. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50–59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

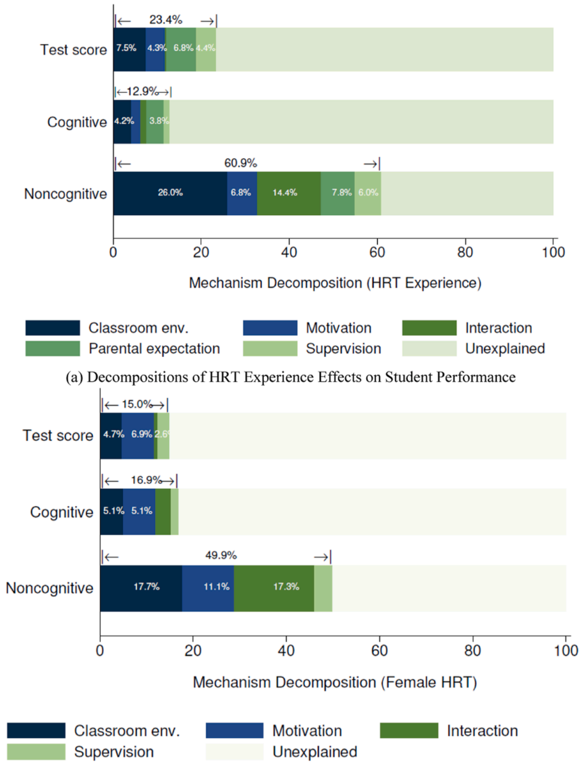
Table 5
Impacts of HRT characteristics on teacher working hours and SBT characteristics.

	Working efforts		SBT characteristics			
	HRT working hours per week (1)	SBT average working hours per week (2)	Exper. (3)	Female (4)	Bachelor+ (5)	Major educ. (6)
HRT exper.	0.243	0.074	0.033	-0.001	-0.006	0.001
	(0.347)	(0.171)	(0.124)	(0.004)	(0.006)	(0.003)
HRT female (= 1)	-0.332	-0.424	1.144	-0.019	-0.009	0.010
	(2.327)	(2.247)	(2.002)	(0.056)	(0.088)	(0.048)
HRT bachelor+ (= 1)	-2.759	2.007	-0.934	0.011	-0.021	-0.011
	(3.350)	(3.783)	(1.628)	(0.076)	(0.102)	(0.027)
HRT major educ. (= 1)	0.683	0.460	-1.277	0.013	-0.072	-0.050
	(5.119)	(2.589)	(2.085)	(0.098)	(0.120)	(0.053)
Observations	12,768	12,507	13,484	13,484	13,484	13,484
R-squared	0.57	0.74	0.69	0.67	0.69	0.68
Mean of Y	51.44	49.85	15.56	0.765	0.505	0.926
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the impacts of HRT characteristics on working hours and SBT characteristics. In columns (1) and (2), the dependent variables are weekly working hours of (1) HRT and (2) SBTs, respectively. In columns (3)–(6), the dependent variables are SBT characteristics, including (3) experience, (4) gender, (5) education level and (6) college major, respectively. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

This study provides novel evidence on the effects of teacher characteristics on a battery of student performance outcomes from the largest developing country, using a random teacher-student assignment sample. Consistent with the studies such as Bloom et al. (2015), Lavy and Boiko (2023), and Mulhern (2020), our findings further deepen the understanding of the education production function by highlighting the importance of teacher management skills that receive less attention. By recognizing the value of effective classroom management, schools can further enhance student performance. One potential avenue is to leverage the roles of teachers with similar responsibilities to HRTs, as they can contribute to creating a conducive learning environment. Incorporating these insights into educational practices (e.g., teacher hiring, training and evaluation) can lead to improved outcomes for students.

The findings in this study open up more research questions. First, our results are for junior high school, and it remains unclear whether our results could be extended to other school stages. Second, it is important to acknowledge that our findings are derived from schools with random teacher-student assignments, which are typically higher-ranked schools. Consequently, a question arises as to whether our policy implications can be generalized to lower-ranked schools and whether this type of assignment represents the best practice across all educational settings. Finally, we did not follow up with the sample students in the long term, and we are unable to



(a) Decompositions of HRT Experience Effects on Student Performance

(b) Decompositions of Female HRT Effects on Student Performance

Fig. 4. Decomposing the HRT Effects by Mechanism.
Notes: Panels (a)–(b) present the decompositions of the HRT effects into mechanisms for each student performance. Panel (a) presents the decompositions of the HRT experience effects into mechanisms of classroom environment, student motivation, parent-teacher interaction, parental expectation and parental supervision. Panel (b) presents the decompositions of the female HRT effects into mechanisms of classroom environment, student motivation, parent-teacher interaction and parental supervision. Each bar represents the corresponding total HRT effects on student performance normalized to 100 percentage points. The percentage of the HRT effects attributable to each mechanism is shown inside each bar.

calculate teacher-specific value-added following Chetty et al. (2014) or find much evidence on long-run effects.

Declaration of Competing Interest

None.

Data availability

The authors do not have permission to share data.

Acknowledgments

We thank participants in numerous seminars for their suggestions. Wei Huang acknowledges research support from the National Natural Science Foundation of China (72373003). Teng Li acknowledges research support from the Youth Program of the National Natural Science Foundation of China (72003202). Yinghao Pan acknowledges research support from the National Natural Science Foundation of China (72003072), the National Social Science Foundation of China (No. 21CJY045), and Social Science Planning Project of Fujian Province (FJ2023MGCA015). All authors contribute equally. All errors are ours.

Appendix

Classroom Assignments in China Junior high school administrations generally assign students in one of the following three ways: 1) pure-random assignment, 2) “snake-shaped” assignment, and 3) “first-last” assignment, as illustrated in Appendix Fig. A1.

Panel (b) in Appendix Fig. A1 shows how pure-random assignment works. Pure-random assignment allocates students into different classes regardless of their pre-entry test scores or other observed student characteristics. Some schools use a computer-aided lottery program, such as Excel or class assignment software, to automatically complete the random assignment. Other schools, especially those in underdeveloped regions, may physically draw lotteries to allocate students into different classes. The whole process is typically open to the public to ensure transparency.

As shown in Panel (c), snake-shaped assignment allocates students from the top ranked to the bottom ranked into different classes following a sequence similar to a snake shape. For better illustration, suppose there are 12 students to be assigned into three classes. All students are indexed based on their rank in terms of pre-entry exam scores. Specifically, Student 1 is the top student, Student 2 ranks second, and so on. The snake-shaped assignment allocates Students 1, 2, and 3 to Classes 1, 2, and 3, respectively. Next, Students 4, 5, and 6 are assigned to Classes 3, 2, and 1, and so on.

Panel (d) illustrates how first-last assignment works. Suppose the same 12 students are to be assigned to three classes. First-last assignment allocates students 1 and 12 into Class 1, Students 2 and 11 to Class 2, and so on. The same procedure is repeated until all students are assigned.

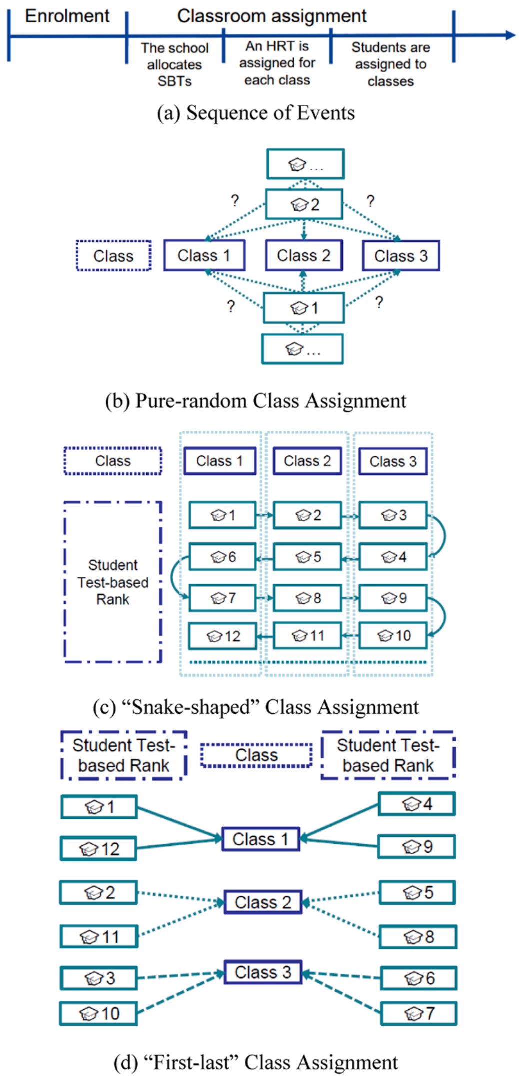


Fig. A1. Random Classroom Assignments.

Notes: Panel (a) explains the sequence of events in the classroom assignment. At stage one, students are enrolled in schools based on their location of residence. At stage two, schools allocate teachers into respective classes and assign one of the SBTs as the HRT. Finally, students are assigned to their classes. Panel (b) introduces the approach that assigns students randomly. Panels (c)-(d) illustrate how snake-shaped assignment and first-last assignment work, respectively.

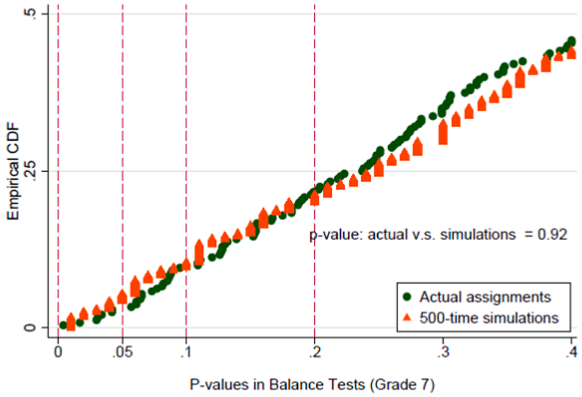


Fig. A2. Empirical CDF of P-values in Balance Tests (Grade 7).

Notes: This Figure compares the cumulative distribution of p-values for the actual assignment and simulated samples for Grade 7. The analysis is parallel to that in Fig. 2.

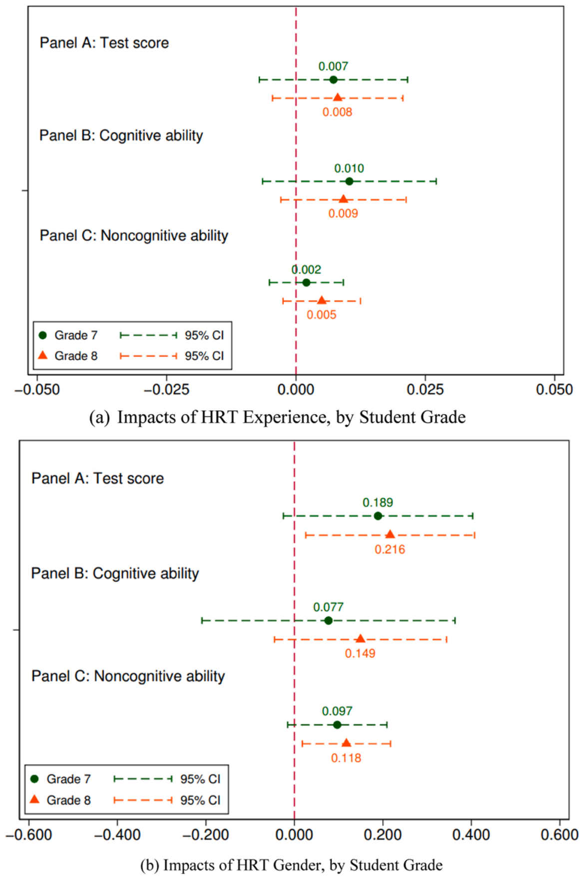
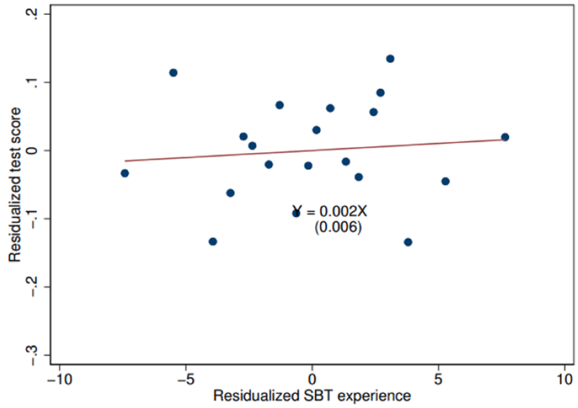
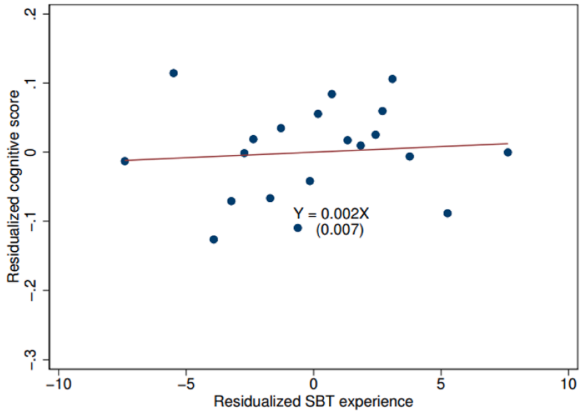


Fig. A3. Impacts of HRT Characteristics in Balanced-Panel Sample.

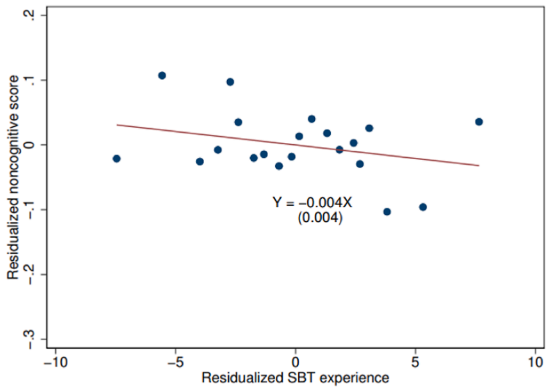
Notes: Panels (a) and (b) present the impacts of HRT characteristics on student performance in the balanced-panel sample, by student grade. The balanced-panel sample used in this study includes only those students who have consistent HRTs across both waves, ensuring that the same HRT-student pairs are tracked.



(a) Test Score



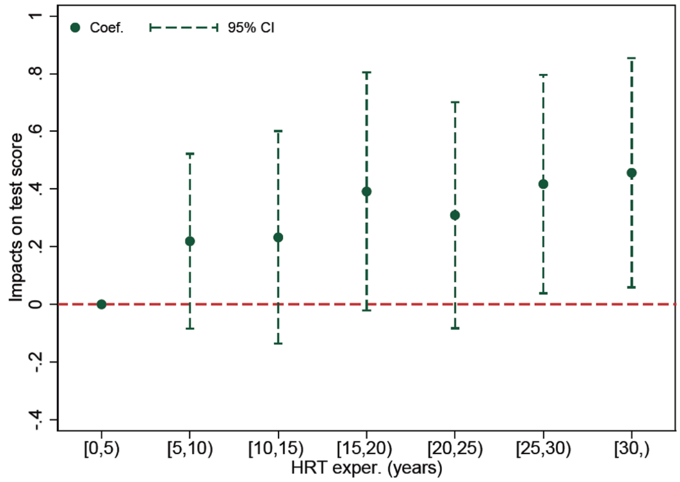
(b) Cognitive Ability



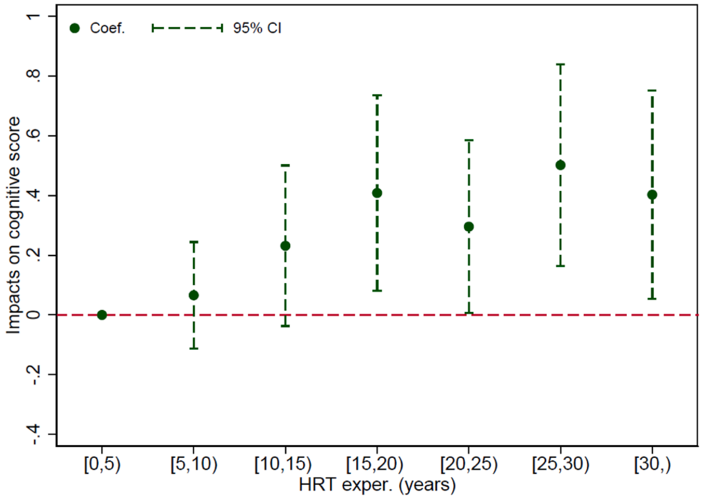
(c) Noncognitive Ability

Fig. A4. Impacts of SBT Experience on Student Performance.

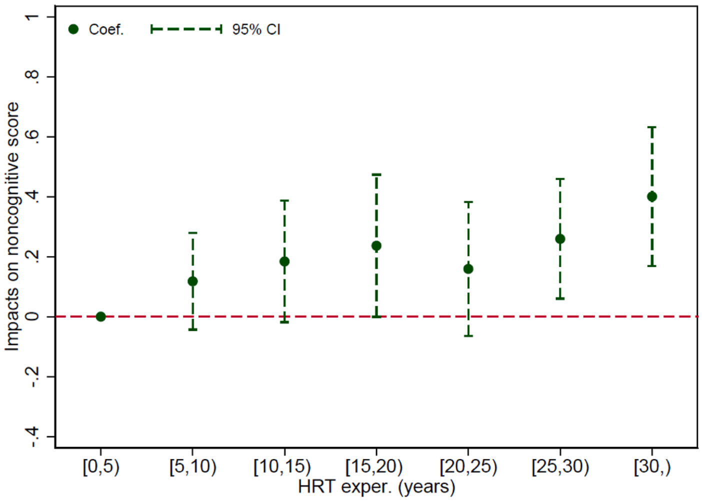
Notes: Panels (a)–(c) present the relation between SBT experience and student performance. We residualize teacher experience and each student outcome with respect to SBT gender, education level, major, class size, students' demographic characteristics, as well as school by grade fixed effects. Each scatter denotes the mean of the student performance against the mean of residualized teacher experience within each of 20 bins. The solid line is for a linear regression fit on the plotted scatters.



(a) Test Score



(b) Cognitive Ability



(c) Noncognitive Ability

(caption on next page)

Fig. A5. Nonlinear Impacts of HRT Experience on Student Performance.
Notes: Panels (a)–(c) present the nonlinear relation between HRT experience and student performance. We divide HRT experience into seven categories, every five years per category. HRTs with experience of less than 5 years are served as the reference group.

Table A1
Determinants of random class assignment.

	Random assignment (= 1) (1)	Random assignment (= 1) (2)
City characteristics		
Eastern city (= 1)	0.12	0.08
	(0.13)	(0.14)
Education (below average as reference)		
Above average (= 1)	0.06	0.14
(0 ~ 1 SD)	(0.12)	(0.13)
Above average (= 1)	0.23*	0.33**
(> 1 SD)	(0.13)	(0.14)
Administrative level (Municities as reference)		
Provincial capital city (= 1)	-0.20	-0.21
	(0.13)	(0.16)
Prefecture (= 1)	-0.05	-0.03
	(0.07)	(0.09)
County (= 1)	-0.00	0.01
	(0.17)	(0.19)
School characteristics		
School ranking (fifth quintile as reference)		
Forth quintile (= 1)		0.57**
		(0.23)
Third quintile (= 1)		0.56***
		(0.18)
Second quintile (= 1)		0.69***
		(0.12)
First quintile (= 1)		0.82***
		(0.17)
Private school (= 1)		0.42***
		(0.10)
Urban (= 1)		-0.05
		(0.08)
Observations	108	107
R-squared	0.055	0.141

Notes: This table reports the relationship between city/ school characteristics and indicators of random class assignment. Schools at the first quintile are top-tier schools in a certain city. Robust standard errors in parentheses are clustered at the city level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2
Worldwide comparison of teachers with similar responsibilities to HRTs.

Country	Appellation	Responsibilities
US	Counselor; Grade-Level Team Leader	Ø Provision of individual and group counseling services to support students in managing personal, social, or emotional challenges Ø Collaboration with teachers, parents, and administrators to identify and address students' academic and behavioral needs Ø Delivery of classroom lessons covering essential topics like conflict resolution, stress management, and healthy decision-making
Canada	Class Advisor; Grade Teacher	Ø Assigned to a specific class or group of students, serving as a mentor, guide, and advocate for their academic and personal growth Ø Assisting students in setting academic and personal goals Ø Monitoring student progress, attendance, and behavior Ø Acting as a liaison between the school and families, communicating with parents about their child's progress and needs
UK	Form Tutor; Form Teacher	Ø Responsible for pastoral care and academic guidance within a specific student group, referred to as a form Ø Supporting students in their form group to achieve academic progress and establish academic goals Ø Offering pastoral care and support to students within the form group Ø Maintaining regular communication with parents or guardians
France	Professeur Principal; Conseiller d'éducation	Ø Communicating with other teachers Ø Organizing and coordinating parent-teacher conferences
Germany	Klassenlehrer; Jahrgangslehrer	Ø Overseeing the academic progress and well-being of students

(*continued on next page*)

Table A2 (*continued*)

Country	Appellation	Responsibilities
Italy	Tutor; Coordinatore Di Classe	Ø Serving as the primary point of contact for parents and guardians Ø Supporting students' social and emotional development and fostering a positive and inclusive classroom environment Ø Regular communication with parents and guardians
Spain	Tutor; Profesor Tutor	Ø Collaboration with other teachers for effective teaching and student support Ø Responsible for guiding and managing a class of students Ø Typically an experienced teacher Ø Maintains contact with students, parents, and other teachers Ø Ensures that students are able to fulfill their potential
Russia	Klassnyy Rukovoditel	Ø Serves as the main point of contact for students, parents, and other teachers
Korea	Seon-Saeng-Nim	Ø Responsible for guiding and managing students' academic and personal development
Japan	Tannin Kyoshi	Ø Responsible for teaching, managing, and caring for a single class of students Ø Daily class management tasks such as attendance and grading
India	Form Teachers	Ø Responsible for a specific class or "form" of students Ø Main responsibilities include teaching the curriculum, maintaining discipline, and monitoring student progress and welfare Ø Regular communication with parents
Australia	Homegroup Teachers	Ø Responsible for creating a positive and inclusive learning environment Ø Monitor the progress of students and provide regular feedback to parents Ø Organize parent-teacher meetings to discuss academic and social progress of students

Notes: This table provides a comparison of teachers from various countries who have similar responsibilities to Homeroom Teachers (HRTs) in their respective education systems. It highlights the common roles and duties these teachers fulfill in terms of student guidance, academic support, and communication with parents and other teachers. The information presented offers insights into the diverse ways in which different countries structure and allocate responsibilities to educators who serve as key points of contact for students' overall well-being and academic progress.

Table A3
Correlation of teacher characteristics.

	Exper. (1)	Female (2)	Bachelor+ (3)	Major educ. (4)
Panel A. HRT characteristics				
Exper.	1.00			
Female	-0.13***	1.00		
Bachelor+	-0.33***	0.18***	1.00	
Major educ.	0.28***	-0.21***	-0.13***	1.00
Panel B. SBT characteristics				
Exper.	1.00			
Female	-0.22***	1.00		
Bachelor+	-0.35***	-0.01*	1.00	
Major educ.	0.01	0.04***	-0.11***	1.00

Notes: This table reports coefficients of correlation for different teacher characteristics. Panels A and B are for HRTs and SBTs, respectively.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4
Components and factor loadings of PCA in noncognitive ability.

Component Analysis Component	Eigenvalue (1)	Proportion (2)	Factor loadings of first component Variable	Comp1 (3)
Comp1	2.44	0.22	Hobbies	0.16
Comp2	1.36	0.12	Curiosity	0.31
Comp3	1.04	0.09	Fast learning	0.45
Comp4	0.98	0.09	Fast reaction	0.43
Comp5	0.96	0.09	Hardworking	0.31
Comp6	0.90	0.08	Seldom escape	0.11
Comp7	0.87	0.08	Social activities	0.26
Comp8	0.77	0.07	Friendly & easygoing	0.31
Comp9	0.69	0.06	Emotional stability	0.17
Comp10	0.56	0.05	Confidence	0.31
Comp11	0.44	0.04	Perception of Friendly classmates	0.31

Notes: This table reports results of PCA using eleven components in Noncognitive ability. Columns (1)-(2) report eigenvalues, and proportion of variance explained. Column (3) reports the factor loading of each variable for the first component.

Table A5
Balance tests for baseline sample.

Teacher type	HRT		SBT	
Specification	Separate regressions	Single regression	Separate regressions	Single regression
	(1)	(2)	(3)	(4)
Panel A. Dependent variable: Experience				
Age in 2013	0.54	0.45	0.44	0.44
Female (= 1)	0.03**	0.03**	0.13	0.37
Local (= 1)	0.58	0.81	0.22	0.50
Rural hukou (= 1)	0.67	0.72	0.09*	0.08*
Only child (= 1)	0.45	0.30	0.68	0.76
Han (= 1)	0.27	0.28	0.83	0.87
Parent bachelor + (= 1)	0.78	0.69	0.92	0.36
Rich during childhood (= 1)	0.94	0.99	0.86	0.79
Kindergarten attendance (= 1)	0.06*	0.03**	0.33	0.24
Late enrolment in prim sch (= 1)	0.65	0.44	0.86	0.63
Transfer in prim sch (= 1)	0.51	0.41	0.91	0.48
Suspension in prim sch (= 1)	0.27	0.39	0.14	0.39
Grade skipping in prim sch (= 1)	0.92	0.93	0.55	0.57
Repeating in prim sch (= 1)	0.92	0.73	0.28	0.53
Class size above 50 (= 1)	0.11	0.11	0.04**	0.04**
Joint Test	–	0.08*	–	0.43
Panel B. Dependent variable: Female				
Age in 2013	0.06*	0.12	0.92	0.67
Female (= 1)	0.31	0.35	0.96	0.99
Local (= 1)	0.42	0.21	0.53	0.68
Rural hukou (= 1)	0.16	0.19	0.13	0.08*
Only child (= 1)	0.26	0.43	0.50	0.80
Han (= 1)	0.08*	0.07*	0.34	0.32
Parent bachelor + (= 1)	0.57	0.98	0.36	0.15
Rich during childhood (= 1)	0.25	0.32	0.31	0.18
Kindergarten attendance (= 1)	0.57	0.74	0.62	0.65
Late enrolment in prim sch (= 1)	0.75	0.39	0.20	0.32
Transfer in prim sch (= 1)	0.55	0.80	0.59	0.31
Suspension in prim sch (= 1)	0.30	0.40	0.52	0.54
Grade skipping in prim sch (= 1)	0.08*	0.15	0.07*	0.06*
Repeating in prim sch (= 1)	0.13	0.28	0.15	0.09*
Class size above 50 (= 1)	0.67	0.67	0.68	0.68
Joint Test	–	0.30	–	0.07*
Panel C. Dependent variable: Bachelor+				
Age in 2013	0.42	0.49	0.27	0.48
Female (= 1)	0.25	0.25	0.17	0.26
Local (= 1)	0.00***	0.02**	0.20	0.50
Rural hukou (= 1)	0.41	0.79	0.19	0.15
Only child (= 1)	0.25	0.62	0.30	0.26
Han (= 1)	0.67	0.63	0.54	0.54
Parent bachelor + (= 1)	0.17	0.18	0.41	0.07*
Rich during childhood (= 1)	0.45	0.58	0.97	0.79
Kindergarten attendance (= 1)	0.35	0.41	0.22	0.18
Late enrolment in prim sch (= 1)	0.77	0.60	0.13	0.09
Transfer in prim sch (= 1)	0.11	0.28	0.74	0.79
Suspension in prim sch (= 1)	0.81	0.86	0.91	0.56
Grade skipping in prim sch (= 1)	0.87	0.94	0.86	0.93
Repeating in prim sch (= 1)	0.98	0.43	0.24	0.38
Class size above 50 (= 1)	0.30	0.30	0.20	0.20
Joint Test	–	0.16	–	0.15
Panel D. Dependent variable: Major educ.				
Age in 2013	0.24	0.17	0.10*	0.66
Female (= 1)	0.51	0.49	0.55	0.54
Local (= 1)	0.67	0.72	0.53	0.94
Rural hukou (= 1)	0.63	0.59	0.38	0.55
Only child (= 1)	0.59	0.60	0.13	0.21
Han (= 1)	0.65	0.69	0.95	0.89
Parent bachelor + (= 1)	0.59	0.47	0.35	0.46
Rich during childhood (= 1)	0.90	0.94	0.32	0.29
Kindergarten attendance (= 1)	0.48	0.43	0.65	0.80
Late enrolment in prim sch (= 1)	0.58	0.40	0.09*	0.13
Transfer in prim sch (= 1)	0.25	0.24	0.06*	0.15
Suspension in prim sch (= 1)	0.91	0.95	0.92	0.94

(*continued on next page*)

Table A5 (*continued*)

Panel D. Dependent variable: Major educ.				
Grade skipping in prim sch (= 1)	0.27	0.22	0.42	0.19
Repeating in prim sch (= 1)	0.90	0.65	0.26	0.41
Class size above 50 (= 1)	0.98	0.98	0.30	0.30
Joint Test	–	0.37	–	0.51

Notes: This table reports results of balance tests in the baseline sample. In Panels A through D, the dependent variables are, respectively, experience in years, gender (female = 1), education (bachelor's degree or above = 1), and college major (education = 1). The independent variables student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes = 1), and parents with bachelor's degree (yes = 1), a set of dummies for the categories of class size (i.e., below 50 and more than 50), as well as school by grade fixed effects. In columns (1) and (3), each cell reports the p-value of the corresponding estimate in separate regressions. In columns (2) and (4), each column represents a single regression, and p-values at the bottom show the joint significance of all independent variables. Robust standard errors are clustered at the school level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6

Balance Tests for Sample of Grade 7.

Teacher type	HRT		SBT	
Specification	Separate regressions (1)	Single regression (2)	Separate regressions (3)	Single regression (4)
Panel A. Dependent variable: Experience				
Age in 2013	0.56	0.47	0.44	0.45
Female (=1)	0.03**	0.03**	0.13	0.35
Local (=1)	0.52	0.75	0.22	0.52
Rural hukou (=1)	0.66	0.69	0.09*	0.08*
Only child (=1)	0.75	0.52	0.54	0.51
Han (=1)	0.26	0.27	0.83	0.87
Parent bachelor + (=1)	0.77	0.73	0.92	0.33
Rich during childhood (=1)	0.93	0.99	0.86	0.80
Kindergarten attendance (=1)	0.07*	0.03**	0.33	0.24
Late enrolment in prim sch (=1)	0.69	0.47	0.86	0.63
Transfer in prim sch (=1)	0.57	0.47	0.91	0.48
Suspension in prim sch (=1)	0.27	0.40	0.14	0.38
Repeating in prim sch (=1)	0.96	0.98	0.55	0.56
Grade skipping in prim sch (=1)	0.91	0.74	0.28	0.54
Class size above 50 (=1)	0.11	0.11	0.04**	0.04**
Joint Test	–	0.07*	–	0.42
Panel B. Dependent variable: Female				
Age in 2013	0.06*	0.12	0.92	0.65
Female (=1)	0.31	0.36	0.96	0.95
Local (=1)	0.42	0.21	0.53	0.73
Rural hukou (=1)	0.16	0.18	0.13	0.09*
Only child (=1)	0.33	0.53	0.27	0.38
Han (=1)	0.08*	0.07*	0.34	0.32
Parent bachelor + (=1)	0.57	0.98	0.36	0.13
Rich during childhood (=1)	0.25	0.32	0.31	0.19
Kindergarten attendance (=1)	0.57	0.74	0.62	0.66
Late enrolment in prim sch (=1)	0.75	0.40	0.20	0.33
Transfer in prim sch (=1)	0.55	0.81	0.59	0.31
Suspension in prim sch (=1)	0.30	0.40	0.52	0.55
Grade skipping in prim sch (=1)	0.08*	0.15	0.07*	0.06*
Repeating in prim sch (=1)	0.13	0.28	0.15	0.09*
Class size above 50 (=1)	0.67	0.67	0.68	0.67
Joint Test	–	0.34	–	0.09*
Panel C. Dependent variable: Bachelor+				
Age in 2013	0.42	0.50	0.27	0.49
Female (=1)	0.25	0.24	0.18	0.25
Local (=1)	0.00***	0.02**	0.20	0.54
Rural hukou (=1)	0.41	0.81	0.19	0.15
Only child (=1)	0.18	0.49	0.28	0.21
Han (=1)	0.67	0.62	0.54	0.54
Parent bachelor + (=1)	0.17	0.19	0.41	0.06*
Rich during childhood (=1)	0.45	0.56	0.97	0.81
Kindergarten attendance (=1)	0.35	0.42	0.22	0.18
Late enrolment in prim sch (=1)	0.77	0.60	0.13	0.09*
Transfer in prim sch (=1)	0.11	0.28	0.74	0.78
Suspension in prim sch (=1)	0.81	0.86	0.91	0.57

(*continued on next page*)

Table A6 (*continued*)

Panel C. Dependent variable: Bachelor+				
Repeating in prim sch (=1)	0.87	0.95	0.86	0.94
Grade skipping in prim sch (=1)	0.98	0.42	0.24	0.39
Class size above 50 (=1)	0.30	0.30	0.20	0.20
Joint Test	–	0.15	–	0.16
Panel D. Dependent variable: Major educ.				
Age in 2013	0.24	0.17	0.10*	0.68
Female (=1)	0.51	0.47	0.55	0.55
Local (=1)	0.67	0.74	0.53	0.92
Rural hukou (=1)	0.63	0.57	0.38	0.55
Only child (=1)	0.76	0.80	0.15	0.28
Han (=1)	0.65	0.69	0.95	0.88
Parent bachelor + (=1)	0.59	0.49	0.35	0.47
Rich during childhood (=1)	0.90	0.94	0.32	0.29
Kindergarten attendance (=1)	0.48	0.43	0.65	0.80
Late enrolment in prim sch (=1)	0.58	0.40	0.09*	0.13
Transfer in prim sch (=1)	0.25	0.25	0.06*	0.15
Suspension in prim sch (=1)	0.91	0.95	0.92	0.94
Grade skipping in prim sch (=1)	0.27	0.22	0.42	0.19
Repeating in prim sch (=1)	0.90	0.66	0.26	0.41
Class size above 50 (=1)	0.98	0.98	0.30	0.30
Joint Test	–	0.40	–	0.52

Notes: This table reports the results of balance tests in the randomly assigned sample in Grade 7. The methodology is the same as that in Appendix Table A5.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7
Balance tests for non-randomly assigned students.

Teacher type	HRT		SBT	
Specification	Separate regressions	Single regression	Separate regressions	Single regression
	(1)	(2)	(3)	(4)
Panel A. Dependent variable: Experience				
Age in 2013	0.48	0.42	0.22	0.25
Female (=1)	0.87	0.93	0.38	0.47
Local (=1)	0.77	0.35	0.13	0.13
Rural hukou (=1)	0.62	0.42	0.99	0.79
Only child (=1)	0.06*	0.09*	0.79	0.92
Han (=1)	0.10*	0.08*	0.77	0.77
Parent bachelor + (=1)	0.14	0.12	0.46	0.47
Rich during childhood (=1)	0.56	0.57	0.87	0.81
Kindergarten attendance (=1)	0.94	0.94	0.12	0.07*
Late enrolment in prim sch (=1)	1.00	0.82	0.07*	0.09*
Transfer in prim sch (=1)	0.16	0.06	0.77	0.79
Suspension in prim sch (=1)	0.16	0.19	0.19	0.24
Repeating in prim sch (=1)	0.96	0.35	0.31	0.70
Grade skipping in prim sch (=1)	0.69	0.64	0.71	0.98
Class size above 50 (=1)	0.97	0.96	0.68	0.69
Joint Test	–	0.00***	–	0.00***
Panel B. Dependent variable: Female				
Age in 2013	0.18	0.40	0.54	0.59
Female (=1)	0.13	0.12	0.90	0.41
Local (= 1)	0.66	0.86	0.68	0.08*
Rural hukou (= 1)	0.26	0.27	0.71	0.07*
Only child (= 1)	0.54	0.41	0.00***	0.00***
Han (= 1)	0.43	0.36	0.88	0.67
Parent bachelor + (= 1)	0.40	0.42	0.47	0.89
Rich during childhood (= 1)	0.29	0.56	0.18	0.06*
Kindergarten attendance (= 1)	0.74	0.62	0.90	0.49
Late enrolment in prim sch (= 1)	0.13	0.17	0.07*	0.06*
Transfer in prim sch (= 1)	0.06*	0.07*	0.87	0.83
Suspension in prim sch (= 1)	0.44	0.38	0.71	0.97
Grade skipping in prim sch (= 1)	0.13	0.68	0.58	0.84
Repeating in prim sch (= 1)	0.09*	0.20	0.23	0.32
Class size above 50 (= 1)	0.59	0.58	0.03**	0.03**
Joint Test	–	0.07**	–	0.00***
Panel C. Dependent variable: Bachelor+				
Age in 2013	0.52	0.58	0.01**	0.00***

(*continued on next page*)

Table A7 (*continued*)

Panel C. Dependent variable: Bachelor+				
Female (= 1)	0.50	0.69	0.34	0.61
Local (= 1)	0.11	0.28	0.37	0.09*
Rural hukou (= 1)	0.93	0.75	0.45	0.46
Only child (= 1)	0.09*	0.05*	0.00***	0.00***
Han (= 1)	0.61	0.79	0.37	0.28
Parent bachelor + (= 1)	0.62	0.56	0.44	0.15
Rich during childhood (= 1)	0.65	0.69	0.14	0.13
Kindergarten attendance (= 1)	0.41	0.18	0.91	0.67
Late enrolment in prim sch (= 1)	0.51	0.51	0.04**	0.16
Transfer in prim sch (= 1)	0.64	0.85	0.44	0.79
Suspension in prim sch (= 1)	0.99	0.96	0.13	0.30
Repeating in prim sch (= 1)	0.42	0.84	0.32	0.79
Grade skipping in prim sch (= 1)	0.57	0.60	0.48	0.71
Class size above 50 (= 1)	0.13	0.12	0.83	0.79
Joint Test	–	0.41	–	0.00***
Panel D. Dependent variable: Major educ.				
Age in 2013	0.31	0.31	0.00***	0.00***
Female (= 1)	0.19	0.18	0.09*	0.67
Local (= 1)	0.91	0.64	0.07*	0.01**
Rural hukou (= 1)	0.70	0.90	0.43	0.01**
Only child (=1)	0.77	0.56	0.00***	0.00***
Han (=1)	0.45	0.59	0.49	0.21
Parent bachelor + (= 1)	0.50	0.39	0.88	0.29
Rich during childhood (= 1)	0.07*	0.04**	0.00***	0.00***
Kindergarten attendance (= 1)	0.27	0.56	0.88	0.87
Late enrolment in prim sch (= 1)	0.36	0.11	0.62	0.24
Transfer in prim sch (= 1)	0.93	0.34	0.01**	0.09*
Suspension in prim sch (= 1)	0.34	0.23	0.08*	0.23
Grade skipping in prim sch (= 1)	0.24	0.63	0.17	0.85
Repeating in prim sch (= 1)	0.33	0.22	0.05**	0.13
Class size above 50 (= 1)	0.12	0.12	0.79	0.87
Joint Test	–	0.69	–	0.00***

Notes: This table reports results of balance tests in the non-randomly assigned sample. The methodology is the same as that in Appendix Table A5.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8
Teacher characteristics and student performance (Excluding Other Controls).

	Test score (1)	Cognitive ability (2)	Noncognitive ability (3)
HRT exper.	0.012**	0.013**	0.009**
	(0.006)	(0.005)	(0.004)
HRT female (=1)	0.217**	0.123	0.101**
	(0.089)	(0.089)	(0.049)
HRT bachelor+ (=1)	-0.095	-0.114*	-0.047
	(0.075)	(0.068)	(0.052)
HRT major educ. (=1)	-0.046	-0.103	0.043
	(0.095)	(0.082)	(0.079)
Observations	12,767	12,855	11,550
R-squared	0.2	0.32	0.16
Student controls	No	No	No
School by grade FEs	Yes	Yes	Yes

Notes: This table reports the effects of teacher characteristics on student test scores and cognitive and noncognitive ability scores. The dependent variables are test scores (column 1), cognitive ability score (column 2), and noncognitive ability score (column 3). Academic test scores are the average of three core-subject test scores, standardized with a mean of 0 and an SD of 1. Cognitive and noncognitive ability are standardized with a mean of 0 and an SD of 1. We also control for school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9
Teacher Characteristics and Student Test Scores, by Core Subject.

	Chinese (1)	(2)	Math (3)	(4)	English (5)	(6)
HRT exper.	0.009*	0.011**	0.013***	0.013**	0.007	0.006
	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)

(*continued on next page*)

Table A9 (*continued*)

	Chinese		Math		English	
	(1)	(2)	(3)	(4)	(5)	(6)
HRT female (=1)	0.140*	0.130*	0.191**	0.232***	0.220***	0.180***
	(0.075)	(0.073)	(0.077)	(0.077)	(0.070)	(0.062)
HRT subject (=1)		0.123		0.041		0.172
		(0.134)		(0.221)		(0.127)
HRT exper. *		-0.006		0.006		-0.000
HRT subject (=1)		(0.009)		(0.011)		(0.008)
SBT exper.	0.001	0.002	0.002	0.003	-0.002	-0.003
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
SBT female	0.005	-0.004	0.081	0.039	-0.064	-0.029
	(0.116)	(0.118)	(0.100)	(0.107)	(0.099)	(0.097)
SBT bachelor+	-0.086	-0.101	-0.027	0.018	-0.108	-0.113
	(0.102)	(0.103)	(0.121)	(0.125)	(0.102)	(0.097)
SBT major educ.	0.462**	0.437*	0.209	0.287	0.272	0.245
	(0.231)	(0.235)	(0.235)	(0.234)	(0.218)	(0.204)
Observations	12,790	12,790	12,796	12,796	12,778	12,778
R-squared	0.29	0.29	0.19	0.19	0.30	0.30
Other HRT characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of teacher characteristics on test scores across core subjects. The dependent variables are standardized test scores of Chinese, Math, and English, with a mean of 0 and an SD of 1. Columns (2), (4), and (6) include the interaction between the indicators of the subject taught by HRTs and HRT experience. Control variables include HRT education level and college major, SBT characteristics, student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10
Heterogeneity analysis.

	Test scores	Cognitive ability	Noncognitive ability
	(1)	(2)	(3)
Panel A. By grade			
Grade 7			
HRT exper.	0.011*	0.014**	0.009***
	(0.006)	(0.006)	(0.003)
HRT female (=1)	0.210**	0.094	0.083
	(0.089)	(0.105)	(0.050)
Grade 8			
HRT exper.	0.012**	0.016***	0.007*
	(0.005)	(0.006)	(0.004)
HRT female (=1)	0.210**	0.160*	0.125***
	(0.083)	(0.082)	(0.044)
Panel B. By student gender			
Female			
HRT exper.	0.009*	0.013**	0.007*
	(0.005)	(0.005)	(0.004)
HRT female (=1)	0.173**	0.078	0.075
	(0.075)	(0.081)	(0.049)
Male			
HRT exper.	0.013**	0.016***	0.008**
	(0.006)	(0.006)	(0.004)
HRT female (=1)	0.237**	0.167*	0.130**
	(0.100)	(0.098)	(0.059)
Panel C. By type of residence (Urban v.s. Rural)			
Urban			
HRT exper.	0.014**	0.017**	0.009**
	(0.006)	(0.006)	(0.004)
HRT female (=1)	0.232**	0.153	0.111**
	(0.103)	(0.109)	(0.051)
Rural			
HRT exper.	0.008	0.010	0.002
	(0.010)	(0.012)	(0.006)
HRT female (=1)	0.161	0.057	0.066
	(0.123)	(0.153)	(0.070)

Notes: This table reports the heterogeneous effects of HRT characteristics. Panels A through C display the estimates by student grade, student gender, and rural/ urban location. In columns (1)-(3), the dependent variables are standardized

test score, cognitive ability, and Noncognitive ability, respectively. Control variables include HRT education level and college major, student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11
Teacher Characteristics and Student Performance (Nonrandom Sample).

	Test score (1)	Cognitive ability (2)	Noncognitive ability (3)
HRT exper.	0.018*	0.019**	0.010*
	(0.010)	(0.008)	(0.006)
HRT female (=1)	0.265***	0.133*	0.113**
	(0.089)	(0.079)	(0.053)
HRT bachelor+ (=1)	0.483***	0.426***	0.211***
	(0.118)	(0.091)	(0.078)
HRT major educ. (=1)	-0.218	-0.107	-0.119
	(0.161)	(0.181)	(0.103)
Observations	5,909	5,940	5,343
R-squared	0.29	0.29	0.15
Student controls	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes

Notes: This table reports the effects of teacher characteristics on student test scores and cognitive and noncognitive ability scores in the non-random sample. The dependent variables are test scores (column 1), cognitive ability score (column 2), and noncognitive ability score (column 3). Academic test scores are the average of three core-subject test scores, standardized with a mean of 0 and an SD of 1. Cognitive and noncognitive ability are standardized with a mean of 0 and an SD of 1. We also control for school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A12
Teacher Characteristics and Student Performance (Pure Random Versus "Balanced" Assignments).

	Test score		Cognitive ability		Noncognitive ability	
	Pure (1)	Balanced (2)	Pure (3)	Balanced (4)	Pure (5)	Balanced (6)
HRT exper.	0.015**	0.014	0.016**	0.016	0.008*	0.008
	(0.007)	(0.009)	(0.007)	(0.012)	(0.004)	(0.005)
HRT female (=1)	0.179	0.276**	0.109	0.053	0.138**	0.084
	(0.118)	(0.122)	(0.126)	(0.164)	(0.060)	(0.071)
HRT bachelor+ (=1)	-0.037	-0.183*	-0.131	-0.006	-0.004	-0.079
	(0.086)	(0.094)	(0.090)	(0.124)	(0.059)	(0.063)
HRT major educ. (=1)	-0.018	-0.140	-0.005	-0.246*	0.042	0.047
	(0.124)	(0.113)	(0.056)	(0.127)	(0.118)	(0.081)
Observations	8,689	4,078	8,762	4,093	7,915	3,635
R-squared	0.26	0.32	0.32	0.37	0.17	0.18
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of teacher characteristics on student test scores and cognitive and noncognitive ability scores. The dependent variables are test scores (columns 1-2), cognitive ability score (columns 3-4), and noncognitive ability score (columns 5-6). Academic test scores are the average of three core-subject test scores, standardized with a mean of 0 and an SD of 1. Cognitive and non-cognitive ability are standardized with a mean of 0 and an SD of 1. Columns 1, 3, and 5 report the results for the sample of students under pure random assignment. Columns 2, 4, and 6 report the results for the sample of students under snake-shaped and first-last assignments. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes =1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A13
Teacher Characteristics and Student Performance (Additional Controls).

	Test score (1)	(2)	Cognitive ability (3)	(4)	Noncognitive ability (5)	(6)
HRT exper.	0.011**	0.010**	0.014***	0.013***	0.007**	0.007**
	(0.005)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)
HRT female (=1)	0.205**	0.202***	0.120	0.117	0.096**	0.100**
	(0.079)	(0.073)	(0.081)	(0.076)	(0.040)	(0.039)
HRT bachelor+ (=1)	-0.065	-0.065	-0.089	-0.088	-0.037	-0.041

(*continued on next page*)

Table A13 (*continued*)

	Test score		Cognitive ability		Noncognitive ability	
	(1)	(2)	(3)	(4)	(5)	(6)
	(0.059)	(0.055)	(0.064)	(0.061)	(0.041)	(0.039)
HRT major educ. (=1)	-0.053	-0.040	-0.095	-0.084	0.050	0.046
	(0.080)	(0.079)	(0.071)	(0.080)	(0.072)	(0.072)
SBT exper.		-0.000		-0.000		-0.005*
		(0.005)		(0.006)		(0.003)
SBT female		0.018		0.011		-0.041
		(0.101)		(0.119)		(0.070)
SBT bachelor+		-0.083		-0.068		-0.013
		(0.105)		(0.111)		(0.080)
SBT major educ.		0.344		0.332		0.043
		(0.233)		(0.257)		(0.144)
Observations	12,767	12,767	12,855	12,855	11,550	11,550
R-squared	0.29	0.29	0.35	0.35	0.19	0.19
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Student early-life experience	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the baseline results (i.e., Table 2) with additional controls. The additional control variables include student early-life experience, such as rich during childhood (yes=1), kindergarten attendance (yes=1), late enrolment (yes=1), transfer (yes=1), suspension (yes=1), grade skipping (yes=1), and repeating (yes=1) in primary school.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A14
HRT Characteristics and Noncognitive Abilities (Big Five Measures).

	Openness Hobbies	Curiosity	Fast learning	Fast reaction	Conscientiousness Hardworking	Seldom escape
	(1)	(2)	(3)	(4)	(5)	(6)
HRT exper.	0.152	0.067	0.276**	0.133	0.097	0.039
	(0.098)	(0.073)	(0.107)	(0.084)	(0.126)	(0.029)
HRT female (=1)	0.617	-0.652	1.734	1.511	4.317**	0.399
	(1.277)	(0.682)	(1.538)	(1.069)	(1.809)	(0.516)
HRT bachelor+ (=1)	-0.727	0.510	-1.645	0.738	-2.345	-0.734
	(1.262)	(0.897)	(1.242)	(1.406)	(2.229)	(0.504)
HRT major educ. (=1)	0.866	-1.288	1.257	0.743	5.165**	-0.241
	(1.464)	(0.991)	(2.656)	(2.911)	(2.189)	(0.763)
R-squared	0.06	0.03	0.09	0.07	0.10	0.04
Mean of Y	88.13	91.44	82.11	85.73	46.97	98.18

	Extraversion Social activities	Agreeableness Friendly & easygoing	Neuroticism Emotional stability	Confidence	Perception of Friendly classmates	AES Noncognitive abilities
	(7)	(8)	(9)	(10)	(11)	(12)
HRT exper.	0.411**	0.091	0.027	-0.003	0.155	0.00362**
	(0.172)	(0.092)	(0.099)	(0.096)	(0.108)	(0.00154)
HRT female (=1)	5.555***	2.028	2.745**	3.276***	0.523	0.0513**
	(1.912)	(1.375)	(1.259)	(1.119)	(1.404)	(0.0212)
HRT bachelor+ (=1)	-1.930	-1.585	0.501	-1.287	1.054	-0.0186
	(3.427)	(1.335)	(1.942)	(1.207)	(1.268)	(0.0204)
HRT major educ. (=1)	0.867	0.697	3.214	1.093	-0.087	0.0251
	(1.670)	(2.845)	(2.812)	(1.848)	(2.009)	(0.0330)
Observations	11,550	11,550	11,550	11,550	11,550	11,550
R-squared	0.24	0.05	0.04	0.07	0.05	
Mean of Y	71.16	82.56	84.81	86.42	87.30	
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of HRT characteristics on noncognitive abilities. In columns (1)-(11), the dependent variables are components of noncognitive abilities as described in Footnote 19. All component variables are multiplied by 100, so the results could be interpreted as percentage point change. Column (12) reports the average effect size of all components. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A15
HRT Characteristics and Student Test Scores (Reweighted).

	Language (1)	Math (2)	Test score (reweighted) (3)
HRT exper.	0.010*	0.013**	0.012**
	(0.005)	(0.005)	(0.005)
HRT female (=1)	0.199**	0.194**	0.210**
	(0.084)	(0.077)	(0.083)
HRT bachelor+ (=1)	-0.082	-0.004	-0.044
	(0.056)	(0.072)	(0.064)
HRT major educ. (=1)	-0.064	-0.028	-0.049
	(0.078)	(0.084)	(0.082)
Observations	12,770	12,796	12,767
R-squared	0.31	0.18	0.26
Student controls	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes

Notes: This table reports the effects of HRT characteristics on reweighted test scores. We define the language scores as the average of Engish and Chinese test scores. The reweighted test score is the average of language and math scores, with a mean of 0 and an SD of 1. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A16
HRT Characteristics (Grade 7) and Teacher-student Attrition (Grade 8).

	HRT change (%) (1)	Student attrition (%) (2)	Student switch within a school (%) (3)	Class reorganization (%) (4)	Attrition in teacher-student pairs (%) (5)
HRT exper.	0.106	0.033	-0.443	-0.005	-0.310
	(0.476)	(0.072)	(0.371)	(0.008)	(0.491)
HRT female (=1)	5.439	0.010	-0.257	-0.131	5.061
	(5.717)	(0.676)	(3.574)	(0.142)	(5.112)
HRT bachelor+ (=1)	0.011	0.053	-1.989	0.050	-1.874
	(5.742)	(0.915)	(2.833)	(0.074)	(5.605)
HRT major educ. (=1)	-14.191	0.103	1.821	0.083	-12.185
	(9.130)	(1.257)	(4.019)	(0.081)	(10.021)
Observations	6,742	6,742	6,742	6,742	6,742
R-squared	0.53	0.27	0.57	0.02	0.54
Mean of Y	7.624	8.499	0.0593	6.942	23.12
Student controls	Yes	Yes	Yes	Yes	Yes
School FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of HRT characteristics in Grade 7 on teacher-student attrition in Grade 8. In columns (1)-(4), the dependent variable is a dummy equal to one for students in Grade 8 (1) in classes with HRT changes, (2) switching from the original class within a school, (3) in reorganized classes, and (4) missing from the survey. In column (5), the dependent variable is a dummy for students having at least one of the above experiences. All dependent variables are multiplied by 100, so the results could be interpreted as percentage point change. The independent variables are measured as in the academic year of 2013-2014 (i.e., Grade 7). Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A17
Robustness Excluding Classes with HRT Change and Using Current HRT Characteristics.

	Test score (1)	Cognitive ability (2)	Noncognitive ability (3)
Panel A. Excluding classes with HRT change			
HRT exper.	0.010*	0.013**	0.009**
	(0.006)	(0.005)	(0.004)
HRT female (=1)	0.215**	0.118	0.098**
	(0.088)	(0.091)	(0.047)
HRT bachelor+ (=1)	-0.080	-0.102	-0.021
	(0.062)	(0.066)	(0.046)
HRT major educ. (=1)	-0.077	-0.112	0.056

(*continued on next page*)

Table A17 (*continued*)

	Test score (1)	Cognitive ability (2)	Noncognitive ability (3)
Panel A. Excluding classes with HRT change			
	(0.085)	(0.076)	(0.080)
Panel B. Using current HRT characteristics			
HRT exper.	0.013**	0.014**	0.008**
	(0.005)	(0.005)	(0.003)
HRT female (=1)	0.194**	0.124	0.093**
	(0.080)	(0.084)	(0.040)
HRT bachelor+ (=1)	-0.020	-0.071	-0.028
	(0.052)	(0.064)	(0.043)
HRT major educ. (=1)	-0.074	-0.070	0.085
	(0.079)	(0.069)	(0.074)
Student controls	Yes	Yes	Yes
Student early-life experience	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes

Notes: This table reports the results of robustness checks. Panel A excludes classes with HRT change and Panel B uses current HRT characteristics. Control variables include HRT education level and college major, student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A18

Analysis of variance.

Dependent variable:	Test score (1)	Cognitive ability (2)	Noncognitive ability (3)
Proportion of Variance Explained by HRT Characteristics			
Exper.	2.39%	1.30%	0.71%
Female (=1)	4.69%	0.60%	0.92%
Bachelor+ (=1)	0.24%	0.19%	0.13%
Major educ. (=1)	0.07%	0.12%	0.05%
Total	7.39%	2.21%	1.81%
Proportion of Variance Explained by SBT Characteristics			
Exper.	0.01%	0.00%	0.02%
Female (=1)	0.18%	0.05%	0.01%
Bachelor+ (=1)	1.32%	0.43%	0.05%
Major educ. (=1)	0.06%	0.01%	0.02%
Total	1.57%	0.50%	0.10%
Model total	8.96%	2.71%	1.91%

Notes: This table reports the explanatory power of observed teacher characteristics. Specifically, we first derive HRT effects by regressing each student performance variable on HRT indicators. Then, the analysis of variance (ANOVA) is applied to calculate the explanatory power by dividing the explained variance of each teacher characteristic by total variance. Thus, the number in each cell represents the proportion of variance that could be explained by the corresponding characteristic.

Table A19

Components and factor loadings of PCA in mechanism indices.

Component Analysis Component	Eigenvalue (1)	Proportion (2)	Factor loadings of first component Variable	Comp1 (3)
Classroom environment				
Comp1	1.75	0.35	Good atmosphere	0.25
Comp2	1.06	0.21	Friends study hard	0.68
Comp3	0.95	0.19	Friends study well	0.67
Comp4	0.93	0.19	Students seldom drink or smoke	0.06
Comp5	0.32	0.06	Students seldom go to Internet cafes	0.13
Motivation				
Comp1	1.19	0.59	Bachelor degree or above	0.71
Comp2	0.81	0.41	Live in big cities	0.71
Parent-teacher interaction				
Comp1	1.30	0.19	Teacher-parent communication	0.60
Comp2	1.09	0.16	Parent-teacher communication	0.49
Comp3	1.04	0.15	Harmonious relationship	0.19
Comp4	0.97	0.14	Parent-teacher meeting	0.35

(*continued on next page*)

Table A19 (*continued*)

Component Analysis Component	Eigenvalue (1)	Proportion (2)	Factor loadings of first component Variable	Comp1 (3)
Comp5	0.92	0.13	Parent fearless of communication	0.30
Comp6	0.89	0.13	HRT familiar with parents	0.35
Comp7	0.79	0.11	HRT respected by parents	0.16
Parental expectation				
Comp1	1.12	0.56	Bachelor degree or above	0.71
Comp2	0.88	0.44	Live in big cities	0.71
Parental supervision				
Comp1	1.32	0.44	Test score	0.37
Comp2	0.96	0.32	Dressing	0.64
Comp3	0.73	0.24	Watching TV	0.68

Notes: This table reports results of PCA using components in mechanism indices. Columns (1)-(2) report eigenvalues, and proportion of variance explained. Column (3) reports the factor loading of each variable for the first component.

Table A20
Summary statistics of mechanism components.

	Observations	Mean	Standard deviation
Classroom environment (%)			
(1) Good atmosphere	12,592	80.40	39.7
(1) Friends study hard	12,592	47.81	49.95
(1) Friends study well	12,592	44.37	49.68
(1) Students seldom drink or smoke	12,592	99.79	4.626
(1) Students seldom go to Internet cafes	12,592	95.23	21.32
Motivation (%)			
(1) Expect to obtain a bachelor degree or above	12,232	83.80	36.84
(1) Expect to live in big cities	12,232	91.20	28.34
Parent-teacher interaction (%)			
(1) Teacher-parent communication	12,286	8.32	27.62
(1) Parent-teacher communication	12,286	51.51	49.98
(1) Harmonious relationship	12,286	79.48	40.39
(1) (11) Parent-teacher meeting	12,286	90.01	29.98
(1) Parent fearless of communication	12,286	76.40	42.47
(1) HRT familiar with parents	12,286	28.38	45.09
(1) HRT respected by parents	12,286	97.88	14.42
Parental supervision (%)			
(1) Test score	12,733	72.14	44.83
(1) Dressing	12,733	28.17	44.98
(1) Watching TV	12,733	44.63	49.71
Parental expectation (%)			
(1) Bachelor degree or above	12,232	83.80	36.84
(1) Live in big cities	12,232	91.20	28.34
Teacher Effort			
(1) HRT working hours	12,501	51.27	29.85
(1) SBT average working hours	12,253	49.82	18.97

Notes: This table presents the summary statistics for mechanism components. Components for classroom environment involve dummies equal to one if (1) a student chooses "agree" or "strongly agree" for the statement, "My class is in a good atmosphere."; at least one friend (2) studies hard and (3) studies well; HRTs report that students in the class seldom (4) drink or smoke and (5) go to Internet cafes. Components for student motivation involve dummies equal to one for students expect to (6) obtain a bachelor degree or above, and (7) live in big cities. Components for parent-teacher interaction indices involve dummies equal to one if (8) HRTs frequently contact parents, (9) parents frequently contact HRTs, (10) HRTs are in a harmonious relationship with parents, (11) parents participate in parent-teacher meeting, (12) parents are fearless of communication with HRTs, (13) HRTs are familiar with parents, and (14) HRTs are respected by parents. Components for parental expectation indices involve dummies equal to one if parents expect their children to (15) obtain a bachelor degree or above and (16) live in big cities. Components for parental expectation indices involve dummies equal to one if parents are strict with (17) test score, (18) dressing and (19) watching TV in student daily life. Component (1)-(19) are multiplied by 100, so the results could be interpreted as proportion. Components for teacher efforts are measured by weekly working hours of (20) HRTs and (21) SBTs.

Table A21
HRT characteristics and mechanism indices (alternative measures).

	Classroom environment (1)	Student motivation (2)	Parent-teacher interaction (3)	Parental expectation (4)	Parental supervision (5)
HRT exper.	0.00775**	0.00484**	0.00662**	0.00637***	0.00458***
	(0.00328)	(0.00234)	(0.00306)	(0.00212)	(0.00140)
HRT female (=1)	0.0601	0.0684**	0.0608*	0.0456	0.0449**

(*continued on next page*)

Table A21 (*continued*)

	Classroom environment (1)	Student motivation (2)	Parent-teacher interaction (3)	Parental expectation (4)	Parental supervision (5)
	(0.0590)	(0.0312)	(0.0345)	(0.0309)	(0.0200)
HRT bachelor+ (=1)	-0.0253	-0.0330	-0.0470	-0.0169	0.0265
	(0.0542)	(0.0337)	(0.0474)	(0.0370)	(0.0327)
HRT major educ. (=1)	-0.163	0.0233	-0.144*	0.0290	0.0335
	(0.101)	(0.0412)	(0.0767)	(0.0379)	(0.0325)
Observations	12,592	12,232	12,286	12,518	12,733
Student controls	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the average effect size of HRT characteristics on mechanism indices following Kling et al. (2007). In columns (1)-(5), the dependent variables are summary indices of (1) classroom environment, (2) student motivation, (3) parent-teacher interaction, (4) parental expectation, and (5) parental supervision. The summary index is the simple average across standardized z-score measures of each component, with a mean of 0 and standard deviation of 1. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A22
Impacts of HRT characteristics on teacher working hours and SBT characteristics, HRT level.

	Working efforts		SBT characteristics			
	HRT working hours per week (1)	SBT average working hours per week (2)	Exper. (3)	Female (4)	Bachelor+ (5)	Major educ. (6)
HRT exper.	0.201	0.086	0.055	-0.002	-0.005	0.000
	(0.333)	(0.185)	(0.114)	(0.004)	(0.005)	(0.003)
HRT female (=1)	0.237	-0.396	1.179	-0.020	-0.005	0.001
	(2.371)	(2.335)	(1.869)	(0.059)	(0.087)	(0.047)
HRT bachelor+ (=1)	-3.785	2.059	0.141	0.035	-0.041	-0.011
	(3.508)	(3.750)	(1.448)	(0.073)	(0.096)	(0.028)
HRT major educ. (=1)	0.239	-0.154	-1.011	0.038	-0.108	-0.032
	(4.740)	(2.452)	(1.862)	(0.101)	(0.120)	(0.062)
Observations	284	278	288	288	288	288
R-squared	0.69	0.73	0.69	0.64	0.69	0.66
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the impacts of HRT characteristics on working hours and SBT characteristics at the HRT level. In columns (1)-(2), the dependent variables are weekly working hours of (1) HRT and (2) SBTs, respectively. In columns (3)-(6), the dependent variables are average SBT characteristics, including (3) experience, (4) gender, (5) education level and (6) college major, respectively. Control variables include a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A23
HRT characteristics and student performance (Inclusion of Mechanism Indices).

	Test score (1)	Cognitive ability (2)	Noncognitive ability (3)	Test score (4)	Cognitive ability (5)	Noncognitive ability (6)
HRT exper.	0.010**	0.013**	0.003	0.010**	0.014***	0.003
	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)	(0.003)
HRT female (=1)	0.186***	0.109	0.053	0.187***	0.110	0.054
	(0.068)	(0.079)	(0.035)	(0.069)	(0.079)	(0.036)
Classroom environment	0.102***	0.066***	0.190***	0.104***	0.068***	0.192***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)
Student motivation	0.152***	0.087***	0.129***	0.220***	0.132***	0.171***
	(0.015)	(0.012)	(0.014)	(0.016)	(0.013)	(0.011)
Parent-teacher interaction	0.005	0.023**	0.106***	0.010	0.025**	0.109***
	(0.012)	(0.011)	(0.012)	(0.011)	(0.010)	(0.012)
Parental supervision	0.089***	0.033***	0.065***	0.093***	0.035***	0.067***
	(0.008)	(0.007)	(0.009)	(0.008)	(0.007)	(0.009)
Parental expectation	0.122***	0.081***	0.076***			
	(0.013)	(0.013)	(0.014)			
Observations	11,299	11,299	11,299	11,299	11,299	11,299
R-squared	0.36	0.36	0.26	0.35	0.36	0.26
Other HRT characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School by grade FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the effects of mechanism indices on student performance. The key independent variables are single-dimensional measures using PCA to consolidate the mechanism components. Each mechanism index is normalized with a mean of 0 and a standard deviation of 1. Control variables include student characteristics, such as indicators of birth cohorts, gender (female= 1), local residence (yes= 1), rural Hukou (yes= 1), being the only child (yes= 1), Han ethnicity (yes= 1), and parents with bachelor's degree (yes= 1), a set of dummies for the categories of class size (i.e., below 30, 30-49, 50-59, and more than 60), as well as school by grade fixed effects. Robust standard errors in parentheses are clustered at the school level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

References

Aaronson, Daniel, Barrow, Lisa, Sander, William, 2007. Teachers and student achievement in the Chicago public high schools. J. Labor Econ. 25 (1), 95–135.
Andersen, Simon Calmar, Beuchert, Louise, Nielsen, Helena Skyt, Thomsen, Mette Kjærgaard, 2020. The effect of teacher's aides in the classroom: evidence from a randomized trial. J. Eur. Econ. Assoc. 18 (1), 469–505.
Angrist, Joshua D, Pathak, Parag A, Walters, Christopher R, 2013. Explaining charter school effectiveness. Am. Econ. J.: Appl. Econ. 5 (4), 1–27.
Antecol, Heather, Eren, Ozkan, Ozbeklik, Serkan, 2015. The effect of teacher gender on student achievement in primary school. J. Labor. Econ. 33 (1), 63–89.
Avvisati, Francesco, Gurgand, Marc, Guyon, Nina, Maurin, Eric, 2014. Getting parents involved: a field experiment in deprived schools. Rev. Econ. Stud. 81 (1), 57–83.
Bau, Natalie, Das, Jishnu, 2020. Teacher value added in a low-income country. Am. Econ. J.: Econ. Pol. 12 (1), 62–96.
Bloom, Nicholas, Lemos, Renata, Sadun, Raffaella, Van Reenen, John, 2015. Does management matter in schools? The Econ. J. 125 (584), 647–674.
Borghans, Lex, Duckworth, Angela Lee, Heckman, James J., ter Weel, Bas, 2008. The economics and psychology of personality traits. J. Hum. Resour. 43 (4), 972–1059.
Borghans, Lex, Golsteyn, Bart HH, Heckman, James J, Humphries, John Eric, 2016. What grades and achievement tests measure. Proc. Natl. Acad. Sci. 113 (47), 13354–13359.
Cameron, Stephen V, Heckman, James J, 2001. The dynamics of educational attainment for black, hispanic, and white males. J. Polit. Econ. 109 (3), 455–499.
Case, Anne, Paxson, Christina, 2008. Stature and status: height, ability, and labor market outcomes. J. Polit. Econ. 116 (3), 499–532.
Chan, Jimmy, Fang, Xian, Wang, Zhi, Zai, Xianhua, Zhang, Qinghua, 2020. Valuing primary schools in urban China. J Urban Econ 115, 103183.
Chetty, Raj, Friedman, John N., Rockoff, Jonah E., 2014. Measuring the impacts of teachers ii: teacher value-added and student outcomes in adulthood. Am. Econ. Rev. 104 (9), 2633–2679.
Chetty, Raj, Looney, Adam, Kroft, Kory, 2009. Salience and taxation: theory and evidence. Am. Econ. Rev. 99 (4), 1145–1177.
Cunha, Flávio, Elo, Irma, Culhane, Jennifer, 2020. Maternal subjective expectations about the technology of skill formation predict investments in children one year later. J. Econometrics.
Deming, David J., 2017. The growing importance of social skills in the labor market. Q. J. Econ. 132 (4), 1593–1640.
Dizon-Ross, Rebecca., 2019. Parents' beliefs about their children's academic ability: implications for educational investments. Am. Econ. Rev. 109 (8), 2728–2765.
Evans, David K, Popova, Anna, 2016. What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. World Bank Res. Obs. 31 (2), 242–270.
Gallego, Francisco A, Malamud, Ofer, Pop-Eleches, Cristian, 2020. Parental monitoring and children's internet use: the role of information, control, and cues. J. Public Econ. 188, 104208.
Gelbach, Jonah B., 2016. When do covariates matter? and which ones, and how much? J. Labor. Econ. 34 (2), 509–543.
Gerritsen, Sander, Plug, Erik, Webbink, Dinand, 2017. Teacher quality and student achievement: evidence from a sample of Dutch twins. J. Appl. Econ. 32 (3), 643–660.
Goldman, Matt, Kaplan, David M, 2018. Comparing distributions by multiple testing across quantiles or Cdf values. J. Econometrics 206 (1), 143–166.
Gong, Jie, Lu, Yi, Song, Hong, 2018. The effect of teacher gender on students' academic and noncognitive outcomes. J. Labor Econ. 36 (3), 743–778.
Hanushek, Eric., 1971. Teacher characteristics and gains in student achievement: estimation using micro data. Am. Econ. Rev. 61 (2), 280–288.
Hanushek, Eric A., 2020. Education production functions. The Economics of Education. Elsevier, pp. 161–170.
Hanushek, Eric A, Rivkin, Steven G, 2012. The distribution of teacher quality and implications for policy. Annu. Rev. Econ. 4 (1), 131–157.
Hanushek, Eric A, Rivkin, Steven G, 2010. Generalizations about using value-added measures of teacher quality. Am. Econ. Rev. 100 (2), 267–271.
Harris, Douglas N, Sass, Tim R, 2011. Teacher training, teacher quality and student achievement. J. Public Econ. 95 (7-8), 798–812.
Heckman, James J, Stixrud, Jora, Urzua, Sergio, 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. J. Labor Econ. 24 (3), 411–482.
Heckman, James, Pinto, Rodrigo, Savelyev, Peter, 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Am. Econ. Rev. 103 (6), 2052–2086.
Hoekstra, Mark, Mouganie, Pierre, Wang, Yaojing, 2018. peer quality and the academic benefits to attending better schools. J. Labor Econ. 36 (4), 841–884.
Islam, Asad., 2019. Parent–teacher meetings and student outcomes: evidence from a developing country. Eur. Econ. Rev. 111, 273–304.
Jackson, C Kirabo, 2018. What do test scores miss? The importance of teacher effects on non–test score outcomes. J. Polit. Econ. 126 (5), 2072–2107.
Jackson, C.Kirabo, Rockoff, Jonah E., Staiger, Douglas O., 2014. Teacher effects and teacher-related policies. Annu. Rev. Econom. 6 (1), 801–825.
Jepsen, Christopher., 2005. Teacher characteristics and student achievement: evidence from teacher surveys. J. Urban Econ. 57 (2), 302–319.
Kane, Thomas J., Taylor, Eric S., Tyler, John H., Wooten, Amy L., 2011. Identifying effective classroom practices using student achievement data. J. Hum. Resour. 46 (3), 587–613.
Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. "Fostering and measuring skills: improving cognitive and non-cognitive skills to promote lifetime success".
Kling, Jeffrey R, Liebman, Jeffrey B, Katz, Lawrence F, 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.
Lavy, Victor., 2020. Expanding school resources and increasing time on task: effects on students' academic and noncognitive outcomes. J. Eur. Econ. Assoc. 18 (1), 232–265.
Lavy, Victor, Daniele Paserman, M, Schlosser, Analia, 2012. Inside the black box of ability peer effects: evidence from variation in the proportion of low achievers in the classroom. The Econ. J. 122 (559), 208–237.
Lavy, Victor, Sand, Edith, 2019. The effect of social networks on students' academic and non-cognitive behavioural outcomes: evidence from conditional random assignment of friends in school. The Econ. J. 129 (617), 439–480.
Malamud, Ofer, Pop-Eleches, Cristian, 2011. Home computer use and the development of human capital. Q. J. Econ. 126 (2), 987–1027.
Mulhern, Christine. 2020. "Beyond teachers: estimating individual guidance counselors' effects on educational attainment," Working paper.
Papay, John P., Kraft, Matthew A., 2015. Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career improvement. J. Public Econ 130, 105–119.
Rockoff, Jonah E., 2004. The impact of individual teachers on student achievement: evidence from panel data. Am. Econ. Rev. 94 (2), 247–252.
Rubin, Donald B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66 (5), 688.

Sass, Tim R, Hannaway, Jane, Xu, Zeyu, Figlio, David N, Feng, Li, 2012. Value added of teachers in high-poverty schools and lower poverty schools. J. Urban Econ. 72 (2-3), 104–122.
Sekhri, Sheetal., 2020. Prestige matters: wage premium and value addition in elite colleges. Am. Econ. J.: Appl. Econ.s 12 (3), 207–225.
Staiger, Douglas O., Rockoff, Jonah E., 2010. Searching for effective teachers with imperfect information. J. Econ. Perspect. 24 (3), 97–118.
Tavares, Priscilla Albuquerque, 2015. The impact of school management practices on educational performance: evidence from public schools in São Paulo. Econ. Educ. Rev. 48, 1–15.
VandenBos, Gary R., 2007. Apa Dictionary of Psychology. American Psychological Association.
Wiswall, Matthew., 2013. The dynamics of teacher quality. J. Public Econ. 100, 61–78.