I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14184

# Teacher Characteristics and Student Performance: Evidence from Random Teacher-Student Assignments in China

Wei Huang
Teng Li
Yinghao Pan
Jinyang Ren

MARCH 2021

DISCUSSION PAPER SERIES

IZA DP No. 14184

# Teacher Characteristics and Student Performance: Evidence from Random Teacher-Student Assignments in China

**Wei Huang**
*National University of Singapore, Emory University and IZA*

**Teng Li**
*Sun Yat-sen University*

**Yinghao Pan**
*National University of Singapore*

**Jinyang Ren**
*Jiangxi University of Finance and Economics*

MARCH 2021



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

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

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

www.iza.org

# ABSTRACT

# Teacher Characteristics and Student Performance: Evidence from Random Teacher-Student Assignments in China*

This paper investigates the impacts of teacher characteristics on student performance using a nationally representative and 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.

**JEL Classification:** I21, J24

**Keywords:** teacher value-added, education production function, student performance

**Corresponding author:**
Wei Huang
National University of Singapore
21 Lower Kent Ridge Rd
Singapore 119077
Singapore

E-mail: huangw@nber.org

* We thank participants in numerous seminars for their suggestions. Wei Huang acknowledges start-up funding support from National University of Singapore (WBS: R-297-000-138-133). Teng Li acknowledges research support from the Youth Program of the National Natural Science Foundation of China (72003202). All errors are ours.

# I. Introduction

Since the seminal work by Hanushek (1971), many economists have been interested in identifying and quantifying the effect of teachers (or teacher value-added) 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 specific factors may contribute to teacher effects (Jepsen, 2005; Aaronson et al., 2007; Bau and Das, 2020). Answers to this question can be used in policy and practice to improve educational outcomes through the processes of recruitment, assignment, and evaluation. For example, understanding how observable characteristics reflect teacher quality provides school administrators helpful information when hiring teachers.

Although there is a long strand of literature addressing this question, the findings are mixed.[1] As mentioned in Rothstein (2010) and Jackson et al. (2014), this inconsistency reveals several common weaknesses in previous studies. First, student-teacher pairs in many contexts are not randomly matched. The findings based on these samples could be driven by unobserved factors such as intended selective sorting in student-teacher pairs (Dieterle et al., 2015). Second, most previous studies use data from schools in specific districts in developed countries such as the US or European countries.[2] Heterogeneous institutions across different regions or countries make the findings in the current literature challenging to generalize, especially for developing countries (Mbiti, 2016). Finally, lacking comprehensive measures on teacher characteristics, student performance, and contemporary information (e.g., parental involvement and teacher effort), most studies focus on one particular teacher characteristic, only a few examine student outcomes beyond test scores, and even fewer explore the underlying mechanisms (Jackson et al., 2014).[3]

In this paper, we first examine the effects of teacher characteristics on student performance, including test scores and cognitive and noncognitive abilities, and then quantify the contributions of

[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.

[2] Krueger (1999) and Mueller (2013) use data from STAR project in Tennessee; Aaronson et al. (2007) collect information from Chicago public schools; and Rockoff (2004) use data from a single New Jersey county.

[3] For example, Rivkin et al. (2005) and Mueller (2013) specifically look at the impact of teacher experience; Antecol et al. (2015) and Gong et al. (2018) focus on the effects of teacher gender. More recent papers such as Jackson et al. (2020) examine the effects on noncognitive abilities. Besides, Gong et al. (2018) further consider the underlying mechanisms.

different mechanisms to student achievement. Our estimates are based on nationally representative data from the China Education Panel Survey (CEPS), which contains rich information on teachers (including homeroom and subject teachers), students, and parents from more than 100 junior high schools across mainland China.[4]

Our data and institutional settings provide an ideal opportunity to estimate the effects of teacher characteristics on student performance. First, students are randomly assigned to classes in most junior high schools in China, with some exceptions.[5] Based on the information reported by school principals and homeroom teachers (HRTs), we restrict our main analysis to the randomly assigned teacher-student sample. Second, the CEPS data are designed to be nationally representative, alleviating concerns regarding external validity. Third, in addition to providing student academic test scores, the CEPS enables us to comprehensively measure student performance in several dimensions, as it additionally collects student cognitive ability scores and related information on noncognitive skills. Fourth, with ample contemporary information on student, parent, and teacher behaviors and subjective opinions collected by the CEPS, we can identify and quantify the underlying mechanisms of teacher effects.

One concern relates to the validity of random assignment in the *reported* randomly assigned teacher-student sample. It is possible that the principals and teachers misreport their method of assigning students. In addition to considering anecdotal evidence,[6] we conduct several statistical tests to alleviate this concern. First, we show that most predetermined student characteristics, such as parental education, only-child indicators, previous academic performance, are insignificantly associated with HRT or SBT 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. In contrast, a parallel analysis for the

[4] The roles played by HRTs and SBTs are detailed in Section 2. In short, in addition to the tasks taken on by subject teachers, HRTs are in charge of student engagement, communicating with parents, and collaborating with subject teachers.

[5] 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.

[6] 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.

comparison between the pseudo- and nonrandom assignments reveals a statistically significant difference in the distribution of the p-values.

The validity of the randomization in teacher-student assignments ensures consistent estimates in our econometric framework. We exploit an OLS regression framework to identify the effects of teacher characteristics, including experience, gender, education level, and college major. Conditional on school-by-grade fixed effects and predetermined student characteristics, our identification hinges on the variations between classes within each school-by-grade cell. Given the different roles of HRTs and SBTs in teaching and classroom management, we examine the characteristics of both HRTs and SBTs.

Our analysis yields three 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 persistent.

In addition, none of the SBT characteristics significantly influences student performance. If teaching skills mattered in the HRT effects, we would expect to observe the effects when an HRT worked as an SBT in other classes. However, our results do not support this. Additionally, the impacts of teacher experience are significant in subjects taught not only by HRTs but also other teachers, with similar magnitudes. If only teaching skills mattered, the impact of an HRT's experience would be more salient in the subject she taught. These results imply that the impacts of HRT characteristics may not be explained by higher teaching skills, which is highlighted in previous literature such as Chetty et al. (2014b); (e.g., Chetty et al., 2014a).

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 other teachers (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 significant evidence shows that teacher effort or better match matters. Likewise, students with a female HRT have a better classroom environment, higher student motivation, and stronger 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. Our findings suggest that managerial practices embedded within HRT characteristics are an important determinant of student performance.

Our findings contribute to several strands of literature. This paper provides new evidence regarding the effects of teacher characteristics on student outcomes from a nationally representative and randomly assigned teacher-student sample from China. Unlike some previous studies that focus on one particular characteristic, we examine a series of teacher characteristics. This helps alleviate concern about the correlation between different characteristics. This exercise also stands out by examining not only test scores but also the cognitive and noncognitive abilities of students (Jackson, 2018). Finally, the evidence builds up the thin literature from the developing world (Evans and Popova, 2016; Mbiti, 2016).

More importantly, this paper contributes to the literature by opening the "black box" of teacher effects on student achievement (Staiger and Rockoff, 2010; Hanushek and Rivkin, 2012; Jackson et al., 2014). Examining the differential effects between HRTs and SBTs and further investigating the mechanisms, we find that managerial practices are critical for teacher value-added. Although there are some exceptions, such as Bloom et al. (2015) and Mulhern (2020),[7] teacher management skills have been understudied in the literature. Our paper takes a step forward by showing that teacher-level management is crucial for student performance.

The paper proceeds as follows. The subsequent section describes junior high schools in China. Section III describes the data. Section IV tests covariate balance in class assignments. Section V presents the results. Section VI discusses the possible mechanisms for the teacher effects. Section VII concludes the paper.

[7] For example, Bloom et al. (2015) find that school-level management is a major determinant of disparities in education quality within and between countries, and Mulhern (2020) shows that high school guidance counselors are almost as important as teachers in the US.

## II. 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".[8] 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).[9] Therefore, the above school-level factors should be controlled for in the analysis, as we will discuss in the next section.

*Class Assignments.* According to the CSLs, junior high schools should not assign students to different ability-based tracks according to their pre-entry scores.[10] 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 Figure 1.

---

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

[9] 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.

[10] 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.

[Insert Figure 1 here]

Panel A in Figure 1 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 B**,** *snake-shaped assignment* allocates students from the top ranked to the bottom ranked into different classes following a sequence similar to a snake shape.[11] 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 C 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.

*Snake-shaped assignment* and *first-last assignment* are also known as balanced assignments. They ensure the average rank of the students is balanced across classes. Both types of balanced assignments are empirically consistent with random assignment (Rubin, 1974; Aaronson et al., 2007). Therefore, following the previous literature such as Gong et al. (2018), we name all the above three assignments as *"random" assignments* for convenience.

In schools applying pure-random or balanced assignments, the school administration and teachers are incentivized to strictly follow the class assignment results except in special cases. 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.[12] 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.

---

[11] In some cities, local education administrations use a hybrid method by combining random assignment with snake-shaped assignment. Chinese news source: http://www.yueqing.gov.cn/art/2020/8/31/art_1443278_55843075.html.

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

However, some junior high schools do assign students to different ability-based classes by tracking.[13] 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 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 or balanced 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.[14] 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 particular 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 the HRT. 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. A typical HRT teaches one of the core subjects and consequently interacts with students more frequently than other SBTs. With assistance from students, HRTs maintain homeroom discipline, offer extracurricular activities, and host weekly homeroom meetings. When necessary, HRTs may also

[13] 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.

[14] 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).

provide students 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 other SBTs to figure out solutions to student problems, especially in relation to academic performance.

The roles of HRTs are not unique to China. Teachers have similar responsibilities in countries such as France, Israel, and Japan. 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 (Avvisati et al., 2014). However, HRTs in other countries may have different roles. 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. Consequently, we should be cautious when generalizing our results to regions or cases with different institutional settings.

## III. 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 nationally representative 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 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. 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

[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).

leave us with 6,742 students in both waves. Table 1 displays the summary statistics for our analysis sample.

[Insert Table 1 here]

***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 might have a different HRT in the second year, as the HRTs may leave their position due to diseases, job turnover or family issues. In the analysis later, we will show that (1) our results are robust to restricting the sample with the same HRTs and using current HRT characteristics, and (2) the teacher effects are persistent in Grade 7 (the first year) and Grade 8 (the following year).

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 the HRTs. Given that there is more than one core SBT for each class, we take the average of their characteristics in our empirical analysis. A typical class has approximately 46 students in our sample.

Unlike some previous studies focusing on single characteristics 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. As shown in Appendix Table A2, female HRTs have fewer years of experience and are more likely to have a college degree. Without including the characteristics in the same equation, it is difficult to conclude that the effect of having a female teacher is from being female itself, having fewer years of experience, or higher education level. Furthermore, it should also be noted that the correlations between the characteristics may differ across different regions or institutional settings.

***Measures of Student Performance.*** We examine 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 both 2013 and 2014. We standardize the scores for each subject with a mean of 70 and an SD of 10 within each school grade. We also measure test scores by taking the average score for the three subjects, with a mean of 70 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).[16] The CEPS conducted a nationwide standardized cognitive ability test assessing students' aptitude in language, visual-spatial abilities, and arithmetic reasoning.[17] The cognitive ability measures are comparable across different schools in each wave, 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 use the Big Five to measure noncognitive skills, similar to the approaches of Borghans et al. (2008) and Heckman et al. (2013).[18] Guided by Jackson (2018) approach, we employ principal component analysis (PCA) to compress the Big Five indicators into a single index (i.e., the first component), with a mean of 0 and an SD of 10. Appendix Table A3 reports the components and factor loadings of the PCA.

---

[16] Although test scores are commonly perceived as cognitive skills in previous literature, we follow the definition 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).

[17] 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.

[18] We 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.

***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. Panel D of Table 1 describes student characteristics. In our sample, students' average age in 2013 is 13. Seven percent of students were born locally, 90 percent are Han ethnicity, and 50 percent are the only child in their families.

## IV. Validity of Randomization

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

*Correlations between Teacher Characteristics and Predetermined Student Characteristics.* Similar to Aaronson et al. (2007), we regress the variables for teacher characteristics on each predetermined student variable while controlling for school-by-grade fixed effects. If the random assignments are appropriately conducted, we expect that the HRT and SBT characteristics will not be associated with the 15 predetermined characteristics. In columns (1) and (3) of Appendix Table A4, we separately regress each variable of HRT teacher characteristics on the 15 predetermined 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 predetermined student variables, using an F-test to show the joint significance of the 15 variables.

With four measures of teacher characteristics (in columns), 15 predetermined student variables (in rows), and two sets of regression equations (separate/joint) for the two types of teachers (HRTs and SBTs) in each panel, we report 240 p-values (4*15*2*2) for teacher characteristics in Appendix Table A4. 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 A5, we restrict the sample to those in Grade 7 for a parallel analysis and find similar results as above. These findings support the reported randomization in class assignments.

*Simulation and Comparison*. Following Chetty et al. (2009), we implement simulations in the spirit of nonparametric permutation tests. If the actual assignments follow the randomization process, the distribution of p-values in the actual assignments should be similar 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 predetermined variable, controlling for school-by-grade fixed effects.

Figure 2(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=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, 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.

[Insert Figure 2 here]

In Appendix Figure A1, we restrict the sample to those in Grade 7, plot the CDF of p-values for actual class assignments, and compare it with that of the 500-times simulation. The patterns in Figure A1 show consistent results with the above.

*Analysis for Schools with Nonrandom Assignment.* We repeat the above analysis for schools with nonrandom assignment. If schools manipulate their class assignments for teachers and students, we expect more significant correlations between teacher and student characteristics. As reported in Appendix Table A6, there are 44 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 Figure 2(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=0.00). This exercise further validates the reliability of the reported random assignments.

# V. 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:

$$Y_{icgs} = \beta_0 + \beta_1 \boldsymbol{HRT_{icgs}} + \beta_2 \boldsymbol{X_{icgs}} + (\beta_3 \boldsymbol{SBT_{icgs}}) + \sigma_{gs} + \varepsilon_{icgs}, \quad (1)$$

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 $\boldsymbol{HRT}_{icgs}$, 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 $\beta_1$, capture the impact of HRT characteristics on student performance. Notably, we use 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 $\boldsymbol{X}_{icgs}$ 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 $\boldsymbol{SBT}_{icgs}$, a vector of variables for SBT characteristics. We control for school-by-grade fixed effects, $\sigma_{gs}$, 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 equation (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 (=0.12*8.18/10); these results are consistent with those in previous studies on returns to teacher experience (e.g., Harris and Sass, 2011). 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.

[Insert Table 2 here]

Column (2) compares the effects of HRT and SBT characteristics on student test scores. None of the SBT characteristics have much predictive power for student test scores, as all coefficients on SBT characteristics are statistically insignificant. 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**

The salient HRT effects and muted SBT effects may be driven by teaching skills. If HRTs impact student test scores with their teaching skills, we should find that the effects of HRTs exist only in their subjects. Table 3 summarizes the effect of HRT characteristics on test scores by subject. As shown in columns (1)-(3), HRT experience significantly impacts Chinese and math test scores. Specifically, a one-SD increase in HRT experience improves Chinese and math test scores by 0.08 SD (=0.10*8.18/10) and 0.11 SD (=0.13*8.18/10), respectively. 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. Moreover, being assigned to female HRTs affects student performance in all subjects.

[Insert Table 3 here]

Next, we investigate whether HRT experience matters only in the subject that HRTs teach. Specifically, we interact HRT experience with an indicator of an HRT's subject. If the HRT's subject drives the effect of HRT experience, we expect the interactions to be positive and significantly different from zero. However, as the estimates in columns (2), (4), and (6) show, the interactions are not statistically significant for 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.

### 5.3 Heterogeneious Impact of Teacher Characteristics

*Grade 7 / Grade 8.* Panels (a) and (b) in Figure 3 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 in the two grades. Table A6 shows the corresponding OLS estimates for Figure 3.

Considering that students are not exactly the same in the two grades,[19] we restrict the analysis to a balanced student sample, and we observe no material changes. We show the results in Appendix Figure A2.

*Student Gender.* Panels (c)-(d) 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)-(f) demonstrate that the HRT effects vary by the type of school location (urban/rural). 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).

[Insert Figure 3 here]

### 5.4 Other Results

*Robustness to Additional Controls and Different Settings.* Appendix Table A8 shows the results when we control for student early-life experiences. The estimates show the results are robust.

Appendix Table A9 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 A10 shows that our results change little if we reweight the test scores across subjects.

*Attrition in HRT-student Pairs*. Changes in the HRT-student pairs after Grade 7 might raise concerns about teacher-student sorting. Specifically, the classroom composition by random assignment in Grade 7 could change in Grade 8 in many ways. Among 6,742 students, 8.5 percent were not

[19] Classroom composition by random assignment in Grade 7 could change in Grade 8 in many ways. Among 6,742 students, 8.5 percent were not followed up with, and only 0.06 percent switched within schools. Approximately 6.9 percent of students experienced class reorganization, and 7.6 percent had a different HRT in Grade 8.

followed up with, and only 0.06 percent switched within schools. Student attrition involves school change (6.3 percent), dropout (1.2 percent), and other reasons (0.9 percent). Approximately 6.9 percent of students experienced class reorganization, and 7.6 percent had a different HRT in Grade 8.

Appendix Table A11 presents the results by regressing these variables on 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 results in Appendix Table A12. In Panel A, we drop the students with a different HRT in the following year and find very consistent results. In Panel B, rather than using the characteristics of initial teachers (i.e., Grade 7), we use the those of current teachers, and our results are consistent. Due to the small number of teacher changes, however, the data do not provide enough statistical power to verify the effects of switching teachers in the sample.

*Nonlinearity*. The relation between teacher experience and student performance (especially test scores) is not necessarily linear (e.g., Rockoff, 2004). Figure 4 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 Figure A3.

[Insert Figure 4 here]

*Analysis of Variance*. An important question raised by previous literature is the extent to which 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. As reported in Appendix Table A13, 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 will further investigate why HRT characteristics matter.

# VI. Mechanisms

## 6.1 Potential Mechanisms

Given that teacher value-added is 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, they may consequently improve 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).[20]

(4) *Teacher efforts and HRT-SBT matching.* As shown by Angrist et al. (2013) and Woessmann (2016), increased instructional time can effectively improve student performance. Additionally, senior or female HRTs may be matched to better SBT teams, possibly improving student performance.

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 10. Appendix Table A14 presents the measurements of each mechanism and the PCA results. Appendix Table A15 summarizes the descriptive statistics for each mechanism component.

[20] 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.

### 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. First, for each mechanism $j$, we replicate equation (1) by replacing $Y_{icgs}$ with the potential mechanism index, $m^j$:

$$m_{icgs}^{j} = \alpha_0^j + \alpha_1^j \boldsymbol{HRT_{icgs}} + \alpha_2^j \boldsymbol{X_{icgs}} + \sigma_{gs}^j + \varepsilon_{icgs}^j. \quad (2)$$

Second, following Heckman et al. (2013), we include the mechanism indices $m_{icgs}^j$ in equation (1):

$$Y_{icgs} = \gamma_0 + \gamma_1 \boldsymbol{HRT_{icgs}} + \sum_{\boldsymbol{j}} \boldsymbol{\varphi^j} \boldsymbol{m_{icgs}^j} + \gamma_2 \boldsymbol{X_{icgs}} + \sigma_{gs} + \varepsilon_{icgs}. \quad (3)$$

Given a specific HRT characteristic *r*, Gelbach (2016) proves that $\hat{\beta}_{1,r} = \hat{\gamma}_{1,r} + \sum_j \hat{\varphi}^j\, \hat{\alpha}_{1,r}^j$. In particular, $r$ takes a value of *exp* or *female* to denote the corresponding coefficient for HRT experience or gender (female =1). The term, $\hat{\varphi}^j \hat{\alpha}_{1,r}^j$, implies the effects of HRT characteristic $r$ on outcomes through mechanism $j$. Therefore, $\hat{\varphi}^j \hat{\alpha}_{1,r}^j / \hat{\beta}_{1,r}$ can be viewed as the proportion of effects of HRT characteristic *r* attributable to mechanism $j$.

*Mechanism Decomposition.* We first follow equation (2) to estimate the impact of HRT characteristics on potential mechanism variables. The results in column (1) of Table 4 imply that a one-SD increase in experience improves the classroom environment by 0.07 SD (=0.09*8.18/10). 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 A16 reports consistent results using an alternative method by Kling et al. (2007).

[Insert Table 4 here]

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.

[Insert Table 5 here]

We further quantify HRT effects by separately calculating the explanatory power of different mechanisms.[21] Panel (a) in Figure 5 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.

[Insert Figure 5 here]

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 Figure 5 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 student noncognitive abilities via more intensive parent-teacher interaction (17.3 percent).

## VII. 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

[21] Following the methodology mentioned above, we estimate equation (3) and report the results in Appendix Table A17. 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.

percent, and 49.9 percent of HRT gender effects on student test scores and cognitive and noncognitive abilities.

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 nationally representative and random teacher-student assignment sample. Consistent with the studies such as Bloom et al. (2015) and Mulhern (2020), our findings further deepen the understanding of the education production function by highlighting the importantce of teacher management skills that receive less attention. The results suggest that management skills merit additional consideration in the design of teacher hiring, performance evaluation, and training programs. Meanwhile, enhancing classroom management by leveraging the roles of teachers with similar duties to HRTs might be another practical approach to improve student performance.

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, our findings are based on schools with random teacher-student assignments, but the question remains whether this type of assignment is best practice. Finally, we did not follow up with the sample students in the long term, and we are unable to find much evidence on long-run effects.

## References

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

**Angrist, Joshua D, Parag A Pathak, and Christopher R Walters.** 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics*, 5(4), 1-27.

**Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik.** 2015. "The Effect of Teacher Gender on Student Achievement in Primary School." *Journal of Labor Economics*, 33(1), 63-89.

**Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin.** 2014. "Getting Parents Involved: A Field Experiment in Deprived Schools." *Review of Economic Studies*, 81(1), 57-83.

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

**Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen.** 2015. "Does Management Matter in Schools?" *The Economic Journal*, 125(584), 647-74.

**Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel.** 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources*, 43(4), 972-1059.

**Borghans, Lex, Bart HH Golsteyn, James J Heckman, and John Eric Humphries.** 2016. "What Grades and Achievement Tests Measure." *Proceedings of the National Academy of Sciences*, 113(47), 13354-59.

**Cameron, Stephen V, and James J Heckman.** 2001. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy*, 109(3), 455-99.

**Case, Anne, and Christina Paxson.** 2008. "Stature and Status: Height, Ability, and Labor Market Outcomes." *Journal of Political Economy*, 116(3), 499-532.

**Chan, Jimmy, Xian Fang, Zhi Wang, Xianhua Zai, and Qinghua Zhang.** 2020. "Valuing Primary Schools in Urban China." *Journal of Urban Economics*, 115, 103183.

**Chetty, Raj, John N. Friedman, and Jonah E. Rockoff.** 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9), 2593-632.

**____.** 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9), 2633-79.

**Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4), 1145-77.

**Cunha, Flávio, Irma Elo, and Jennifer Culhane.** 2020. "Maternal Subjective Expectations About the Technology of Skill Formation Predict Investments in Children One Year Later." *Journal of Econometrics*.

**Deming, David J.** 2017. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics*, 132(4), 1593-640.

**Dieterle, Steven, Cassandra M Guarino, Mark D Reckase, and Jeffrey M Wooldridge.** 2015. "How Do Principals Assign Students to Teachers? Finding Evidence in Administrative Data and the Implications for Value Added." *Journal of Policy Analysis and Management*, 34(1), 32-58.

**Dizon-Ross, Rebecca.** 2019. "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments." *American Economic Review*, 109(8), 2728-65.

**Evans, David K, and Anna Popova.** 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *The World Bank Research Observer*, 31(2), 242-70.

**Gallego, Francisco A, Ofer Malamud, and Cristian Pop-Eleches.** 2020. "Parental Monitoring and Children's Internet Use: The Role of Information, Control, and Cues." *Journal of Public Economics*, 188, 104208.

**Gelbach, Jonah B.** 2016. "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics*, 34(2), 509-43.

**Gerritsen, Sander, Erik Plug, and Dinand Webbink.** 2017. "Teacher Quality and Student Achievement: Evidence from a Sample of Dutch Twins." *Journal of Applied Econometrics*, 32(3), 643-60.

**Goldman, Matt, and David M Kaplan.** 2018. "Comparing Distributions by Multiple Testing across Quantiles or Cdf Values." *Journal of Econometrics*, 206(1), 143-66.

**Gong, Jie, Yi Lu, and Hong Song.** 2018. "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes." *Journal of Labor Economics*, 36(3), 743-78.

**Hanushek, Eric.** 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review*, 61(2), 280-88.

**Hanushek, Eric A.** 2020. "Education Production Functions," *The Economics of Education.* Elsevier, 161-70.

**Hanushek, Eric A, and Steven G Rivkin.** 2012. "The Distribution of Teacher Quality and Implications for Policy." *Annu. Rev. Econ.*, 4(1), 131-57.

**____.** 2010. "Generalizations About Using Value-Added Measures of Teacher Quality." *American Economic Review*, 100(2), 267-71.

**Harris, Douglas N, and Tim R Sass.** 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics*, 95(7-8), 798-812.

**Heckman, James J, Jora Stixrud, and Sergio Urzua.** 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor*

*Economics*, 24(3), 411-82.
**Heckman, James, Rodrigo Pinto, and Peter Savelyev.** 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, 103(6), 2052-86.
**Islam, Asad.** 2019. "Parent–Teacher Meetings and Student Outcomes: Evidence from a Developing Country." *European Economic Review*, 111, 273-304.
**Jackson, C Kirabo.** 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy*, 126(5), 2072-107.
**Jackson, C Kirabo, Shanette C Porter, John Q Easton, Alyssa Blanchard, and Sebastián Kiguel.** 2020. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights*, 2(4), 491-508.
**Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger.** 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics*, 6(1), 801-25.
**Jepsen, Christopher.** 2005. "Teacher Characteristics and Student Achievement: Evidence from Teacher Surveys." *Journal of Urban Economics*, 57(2), 302-19.
**Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz.** 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1), 83-119.
**Krueger, Alan B.** 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 114(2), 497-532.
**Lavy, Victor, M Daniele Paserman, and Analia Schlosser.** 2012. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *The Economic Journal*, 122(559), 208-37.
**Lavy, Victor, and Edith Sand.** 2019. "The Effect of Social Networks on Students' Academic and Non-Cognitive Behavioural Outcomes: Evidence from Conditional Random Assignment of Friends in School." *The Economic Journal*, 129(617), 439-80.
**Malamud, Ofer, and Cristian Pop-Eleches.** 2011. "Home Computer Use and the Development of Human Capital." *Quarterly Journal of Economics*, 126(2), 987-1027.
**Mbiti, Isaac M.** 2016. "The Need for Accountability in Education in Developing Countries." *Journal of Economic Perspectives*, 30(3), 109-32.
**Mueller, Steffen.** 2013. "Teacher Experience and the Class Size Effect—Experimental Evidence." *Journal of Public Economics*, 98, 44-52.
**Mulhern, Christine.** 2020. "Beyond Teachers: Estimating Individual Guidance Counselors' Effects on Educational Attainment," Working paper,
**Papay, John P., and Matthew A. Kraft.** 2015. "Productivity Returns to Experience in the Teacher

Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement." *Journal of Public Economics*, 130, 105-19.

**Rivkin, Steven G, Eric A Hanushek, and John F Kain.** 2005. "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2), 417-58.

**Rockoff, Jonah E.** 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2), 247-52.

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

**Rubin, Donald B.** 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5), 688.

**Sass, Tim R, Jane Hannaway, Zeyu Xu, David N Figlio, and Li Feng.** 2012. "Value Added of Teachers in High-Poverty Schools and Lower Poverty Schools." *Journal of Urban Economics*, 72(2-3), 104-22.

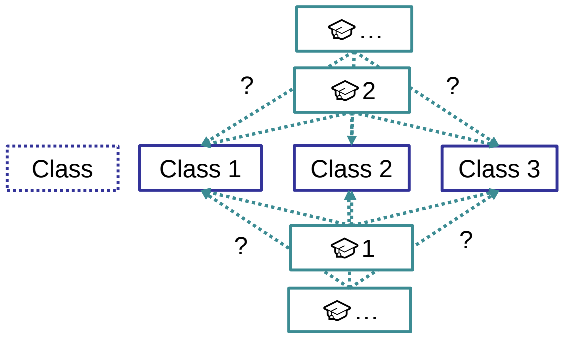
**Sekhri, Sheetal.** 2020. "Prestige Matters: Wage Premium and Value Addition in Elite Colleges." *American Economic Journal: Applied Economics*, 12(3), 207-25.

**Staiger, Douglas O., and Jonah E. Rockoff.** 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives*, 24(3), 97-118.

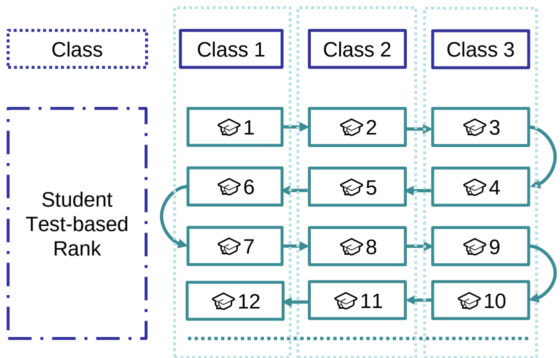
**VandenBos, Gary R.** 2007. "*APA Dictionary of Psychology*." American Psychological Association.

**Wiswall, Matthew.** 2013. "The Dynamics of Teacher Quality." *Journal of Public Economics*, 100, 61-78.

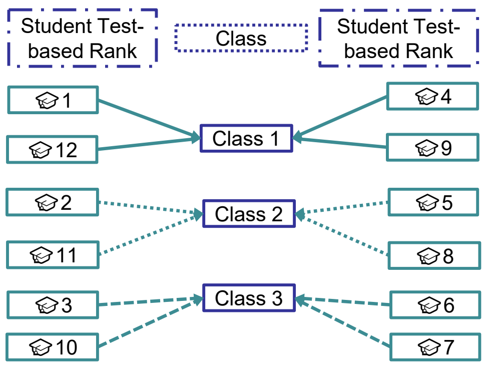
**Woessmann, Ludger.** 2016. "The Importance of School Systems: Evidence from International Differences in Student Achievement." *Journal of Economic Perspectives*, 30(3), 3-32.




(a) Pure-random Class Assignment



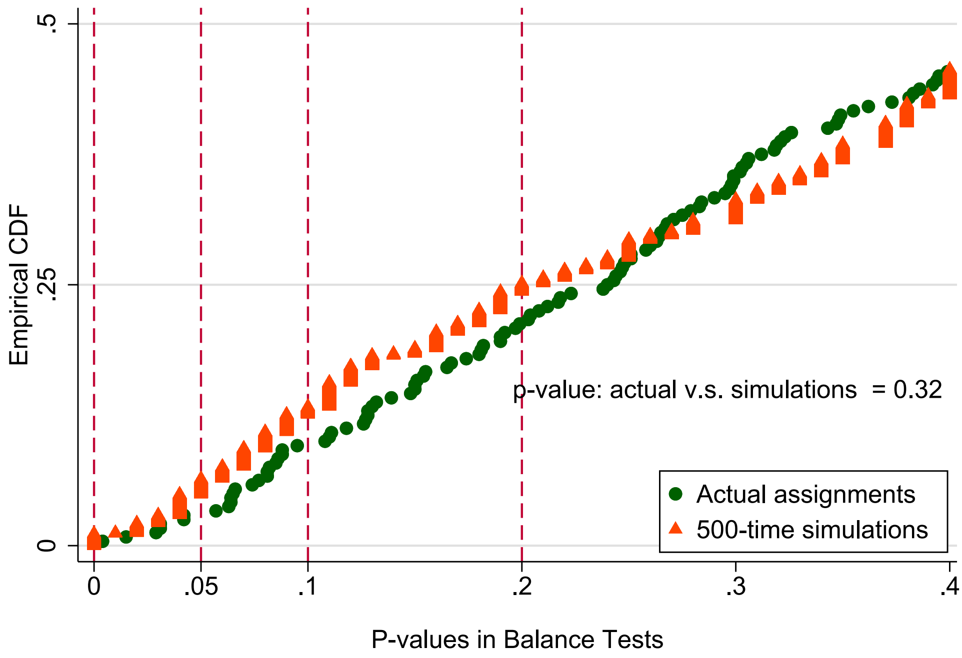

(b) "Snake-shaped" Class Assignment



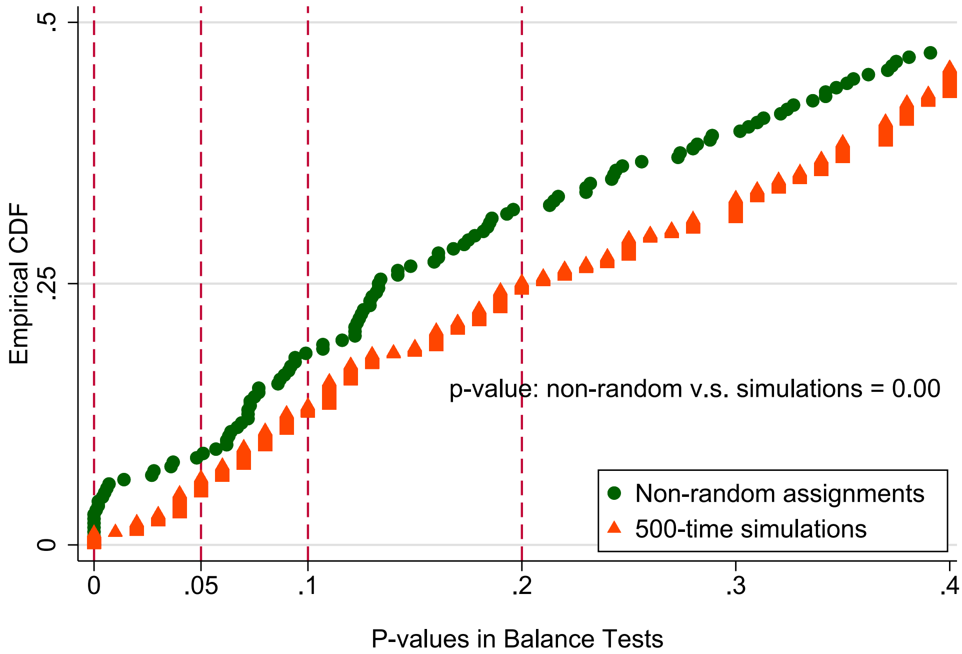

(c) "First-last" Class Assignment

Figure 1: Random Classroom Assignments

*Notes:* Panels (a)-(c) present three common approaches to assign students in junior high schools. Panel (a) introduces the approach that assigns students randomly. Panels (b)-(c) introduce the approaches that assign students into different classes by student rank using the pre-enrolment test scores, including (b) the "snake-shaped" assignment and (c) "first-last" assignment.



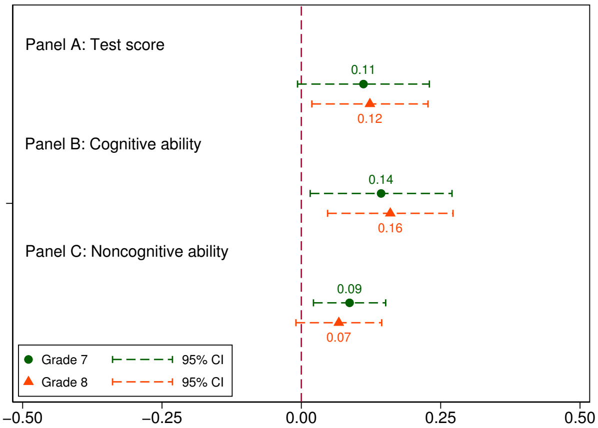

(a) Actual Assignment V.S. 500-time Simulations for the Randomly-assigned Sample



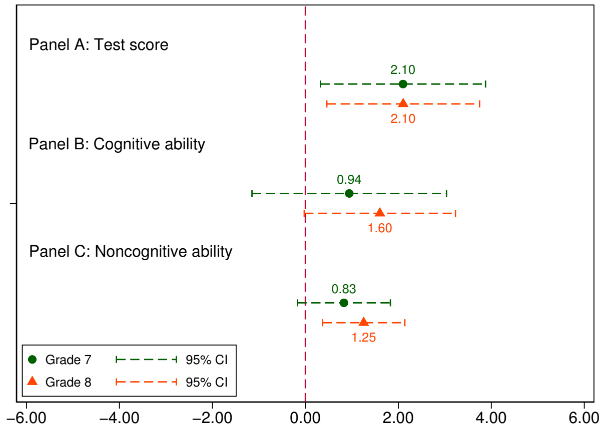

(b) Actual Assignment V.S. 500-time Simulations for Nonrandomly-assigned Sample

Figure 2: Empirical CDF of P-values in Balance Tests

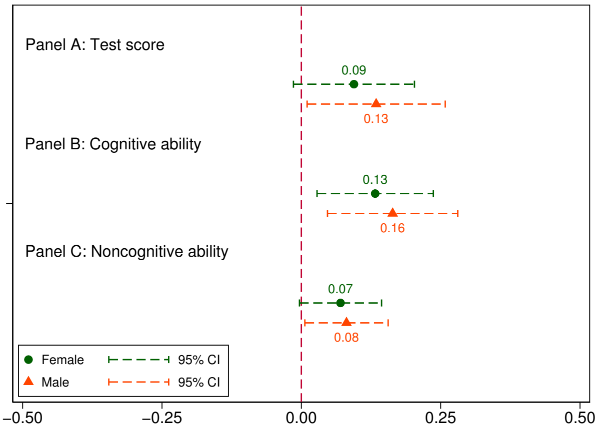
*Notes:* Panels (a)-(b) compare the cumulative distribution of p-vaues for actual assignment and pseudo-samples. Panels (a)-(b) present the comparison for the randomly- and nonrandomly-assigned sample, respectively. For simulations, we randomly assign students to classes within the same grade and school 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.



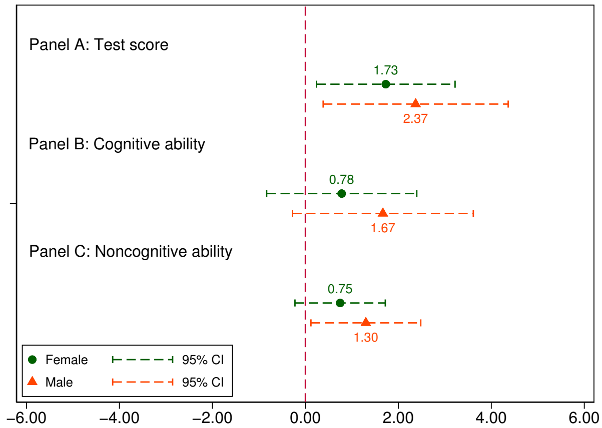

(a) Impacts of HRT experience, by Student Grade



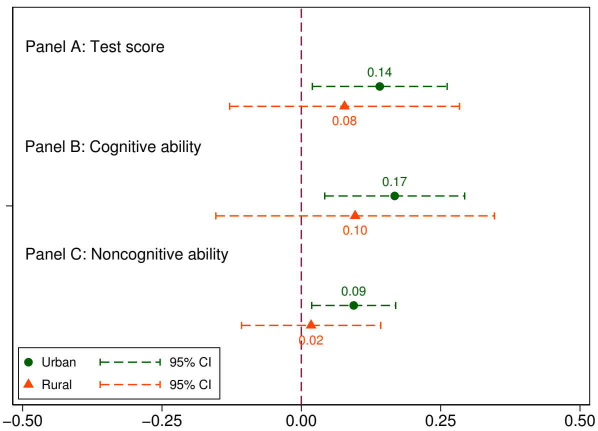

(b) Impacts of HRT HRT Being Female, by Student Grade



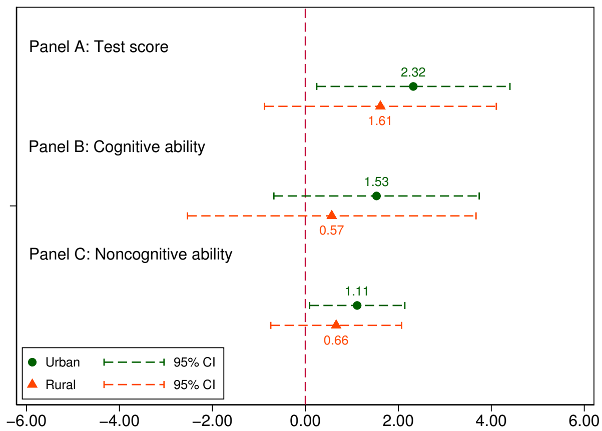

(c) Impacts of HRT Experience, by Student Gender




(d) Impacts of HRT Being Female, by Student Gender



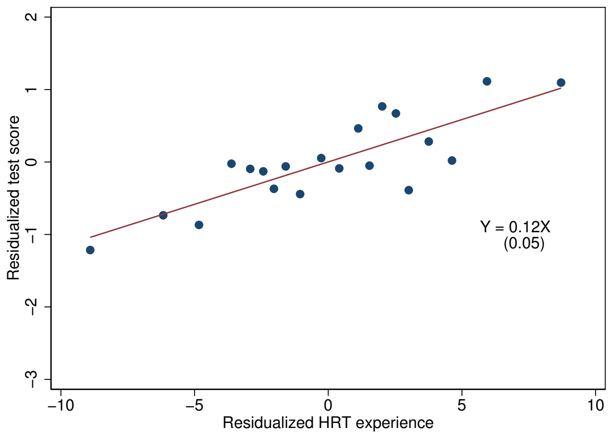

(e) Impacts of HRT Experience, by Type of Location



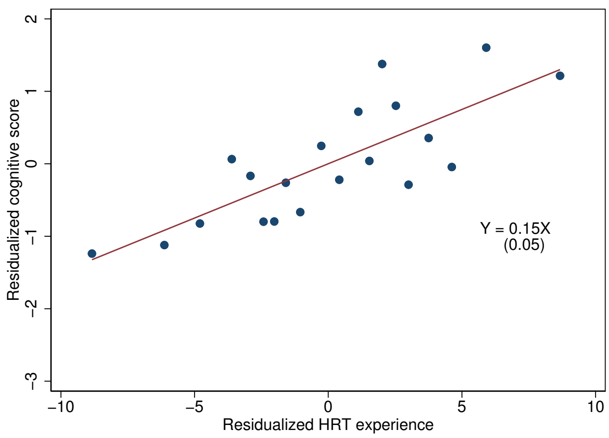

(f) Impacts of HRT Being Female, by Type of Location

Figure 3: Heterogeneous Impacts of HRT Characteristics

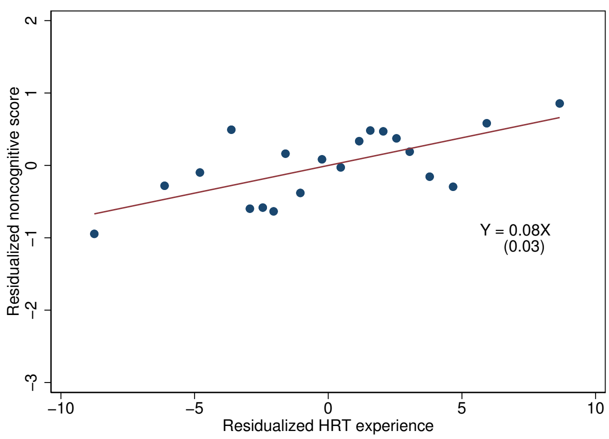
*Notes:* Panels (a)-(f) present the heterogeneous effects of HRT characteristics based on equation (1), by student grade (a-b), student gender (c-d), and type of location (e-f). Control variables include HRT education level and college major, student characteristics, such as indicators of birth cohorts, 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.




(a) Test Score



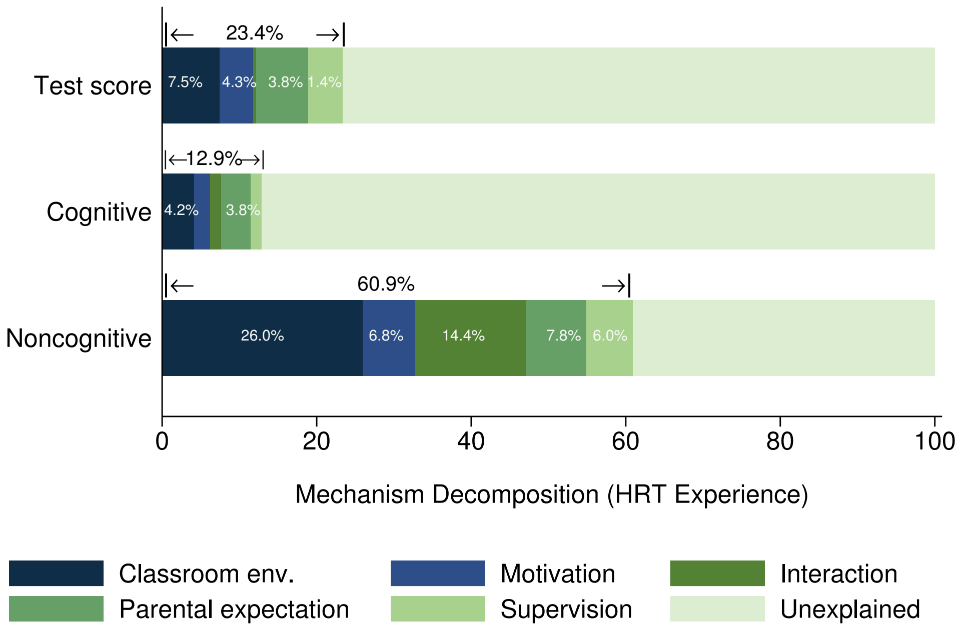

(b) Cognitive Ability



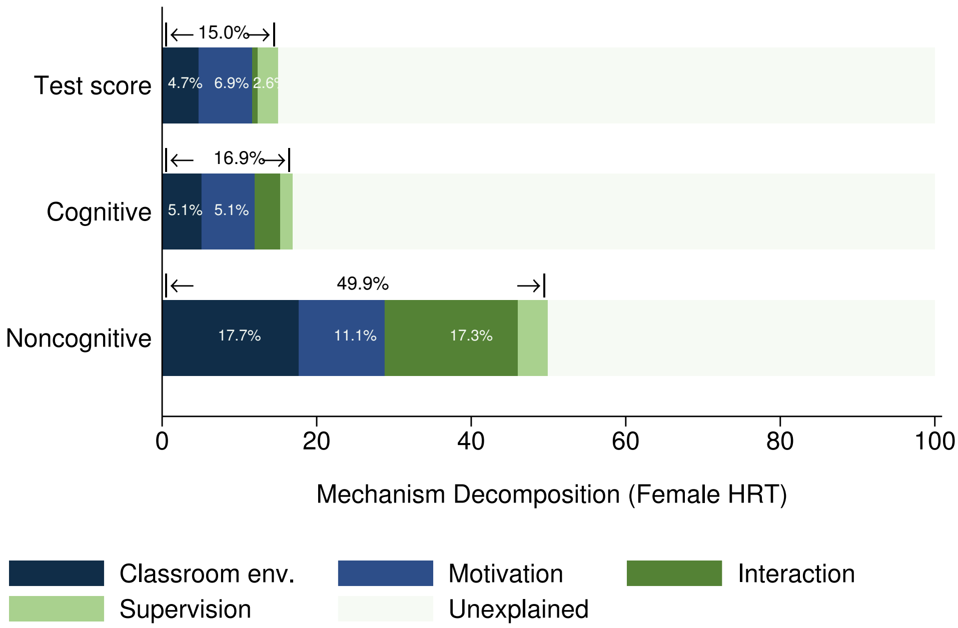

(c) Noncognitive Ability

Figure 4: 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.




(a) Decompositions of HRT Experience Effects on Student Performance




(b) Decompositions of Female HRT Effects on Student Performance

Figure 5: 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.

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 | 70 | 10 |
| Cognitive ability | 12,855 | 0 | 10 |
| Non-cognitive ability | 11,550 | 0 | 10 |
| | | | |
| ***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 70 and an SD of 10. Cognitive and non-cognitive ability are standardized with a mean of 0 and an SD of 10. Past experience is reported by students.

Table 2: Teacher Characteristics and Student Performance

| | Test score | | Cognitive ability | | Non-cognitive ability | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| HRT exper. | 0.12** | 0.11** | 0.15*** | 0.14*** | 0.08** | 0.08** |
| | (0.05) | (0.05) | (0.05) | (0.05) | (0.03) | (0.03) |
| HRT female (=1) | 2.10** | 2.06*** | 1.25 | 1.21 | 1.03** | 1.07** |
| | (0.84) | (0.78) | (0.86) | (0.80) | (0.44) | (0.43) |
| HRT bachelor+ (=1) | -0.58 | -0.56 | -0.84 | -0.82 | -0.33 | -0.37 |
| | (0.61) | (0.56) | (0.66) | (0.62) | (0.43) | (0.41) |
| HRT major educ. (=1) | -0.55 | -0.42 | -0.98 | -0.86 | 0.51 | 0.48 |
| | (0.81) | (0.81) | (0.71) | (0.80) | (0.72) | (0.73) |
| SBT exper. | | 0.00 | | 0.00 | | -0.05 |
| | | (0.05) | | (0.06) | | (0.03) |
| SBT female | | 0.09 | | 0.10 | | -0.37 |
| | | (1.06) | | (1.21) | | (0.72) |
| SBT bachelor+ | | -0.84 | | -0.71 | | -0.16 |
| | | (1.09) | | (1.16) | | (0.84) |
| SBT major educ. | | 3.59 | | 3.43 | | 0.64 |
| | | (2.45) | | (2.70) | | (1.56) |
| 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 1-2), cognitive ability score (columns 3-4), and non-cognitive ability score (columns 5-6). Academic test scores are the average of three core-subject test scores, standardized with a mean of 70 and an SD of 10. Cognitive and non-cognitive ability are standardized with a mean of 0 and an SD of 10. 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 3: HRT Characteristics and Student Test Scores, by Core Subject

| | Chinese | | Math | | English | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| HRT exper. | 0.10* | 0.12** | 0.13** | 0.13** | 0.08 | 0.06 |
| | (0.06) | (0.06) | (0.05) | (0.06) | (0.05) | (0.05) |
| HRT female (=1) | 1.46* | 1.38 | 1.94** | 2.30*** | 2.24*** | 1.80*** |
| | (0.83) | (0.83) | (0.77) | (0.76) | (0.77) | (0.66) |
| HRT subject (=1) | | 1.24 | | 0.01 | | 1.71 |
| | | (1.27) | | (2.10) | | (1.29) |
| HRT exper. * | | -0.06 | | 0.08 | | 0.00 |
| HRT subject (=1) | | (0.08) | | (0.11) | | (0.08) |
| 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 70 and an SD of 10. Columns (1), (3), and (5) include the indicators of the subject taught by HRTs. 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.

Table 4: Impacts of HRT Characteristics on Mechanism Indices

| | Classroom environment | Student motivation | Parent-teacher interaction | Parental expectation | Parental supervision |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| HRT exper. | 0.09** | 0.06** | 0.10* | 0.09*** | 0.06*** |
| | (0.04) | (0.03) | (0.06) | (0.03) | (0.02) |
| HRT female (=1) | 1.04** | 0.89** | 1.86*** | 0.64 | 0.69** |
| | (0.46) | (0.41) | (0.64) | (0.43) | (0.30) |
| HRT bachelor+ (=1) | -0.46 | -0.43 | -1.57 | -0.25 | 0.43 |
| | (0.50) | (0.45) | (1.03) | (0.51) | (0.48) |
| HRT major educ. (=1) | -0.93 | 0.26 | -1.99 | 0.37 | 0.50 |
| | (0.86) | (0.54) | (1.27) | (0.52) | (0.43) |
| 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 10. 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 | SBT average working hours per week | Exper. | Female | Bachelor+ | Major educ. |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| HRT exper. | 0.25 | 0.07 | 0.03 | -0.00 | -0.01 | 0.00 |
| | (0.35) | (0.17) | (0.13) | (0.00) | (0.01) | (0.00) |
| HRT female (=1) | -0.27 | -0.43 | 1.10 | -0.02 | -0.00 | 0.01 |
| | (2.35) | (2.25) | (2.01) | (0.06) | (0.09) | (0.05) |
| HRT bachelor+ (=1) | -2.80 | 2.00 | -0.92 | 0.00 | -0.03 | -0.01 |
| | (3.39) | (3.83) | (1.65) | (0.08) | (0.10) | (0.03) |
| HRT major educ. (=1) | 0.80 | 0.60 | -1.28 | 0.01 | -0.07 | -0.05 |
| | (5.23) | (2.60) | (2.13) | (0.10) | (0.12) | (0.05) |
| Observations | 12,501 | 12,253 | 12,767 | 12,767 | 12,767 | 12,767 |
| R-squared | 0.57 | 0.74 | 0.69 | 0.67 | 0.69 | 0.68 |
| Mean of Y | 51.27 | 49.82 | 15.50 | 0.769 | 0.504 | 0.928 |
| 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)-(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.

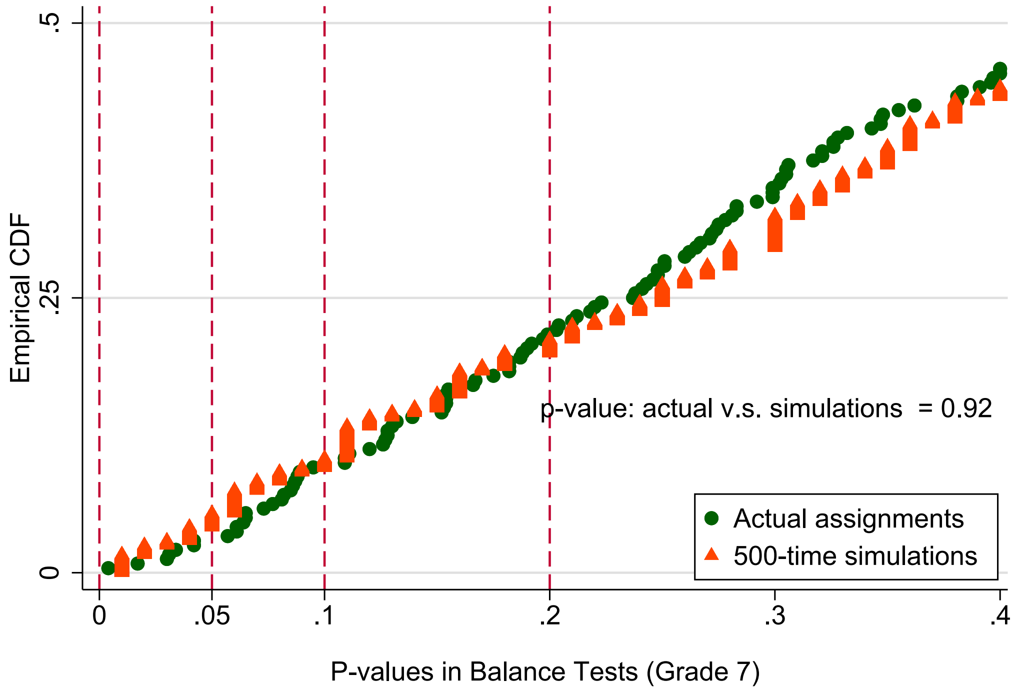
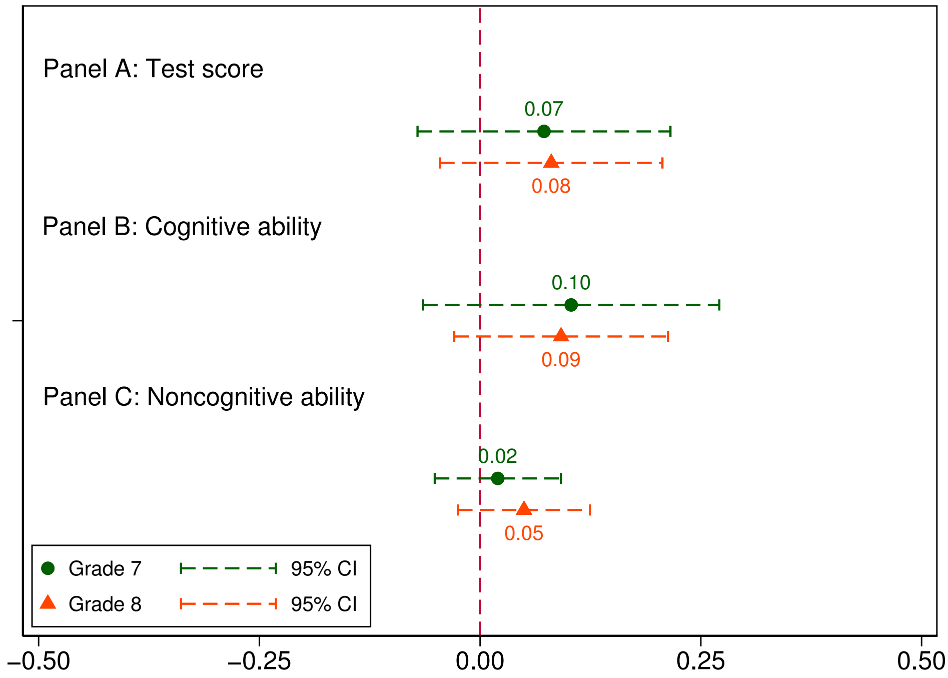
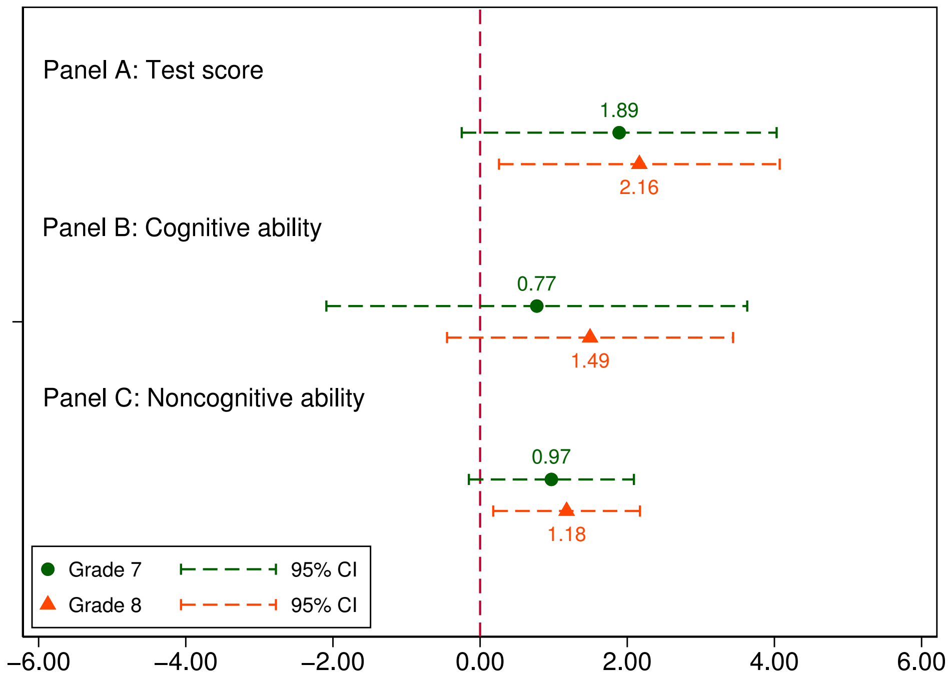



Figure A1: Empirical CDF of P-values in Balance Tests (Grade 7)

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



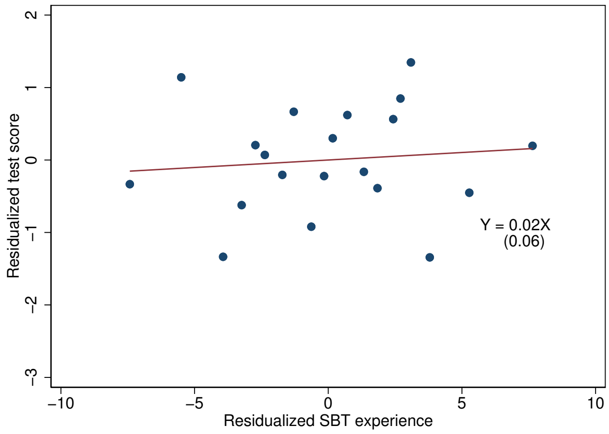

(a) Impacts of HRT Experience, by Student Grade



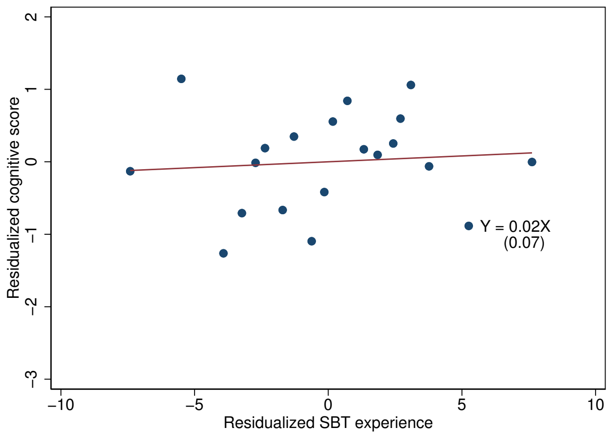

(b) Impacts of HRT Gender, by Student Grade

Figure A2: Impacts of HRT Characteristics in Balanced Sample

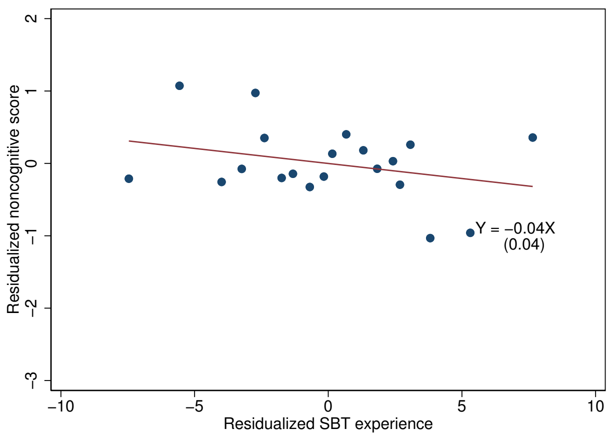
*Notes:* Panels (a)-(b) present the impacts of HRT characteristics on student performance in the balanced sample, by student grade.




(a) Test Score




(b) Cognitive Ability




(c) Noncognitive Ability

Figure A3: 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.

Table A1: Determinants of Random Class Assignment

| | Random assignment (=1) | Random assignment (=1) |
|---|---|---|
| | (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: 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 A3: Components and Factor Loadings of PCA in Noncognitive Ability

| Component Analysis | | | Factor loadings of first component | |
|---|---|---|---|---|
| Component | Eigenvalue | Proportion | Variable | Comp1 |
| | (1) | (2) | | (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 A4: 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* |

Table A4 (Continued)

| Teacher type | HRT | | SBT | |
|---|---|---|---|---|
| Specification | Separate regressions | Single regression | Separate regressions | Single regression |
| | (1) | (2) | (3) | (4) |
| ***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 |
| 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 A5: Balance Tests for Sample of Grade 7

| 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 | 0.56 | 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* |

Table A5 (Continued)

| Teacher type | HRT | | SBT | |
|---|---|---|---|---|
| Specification | Separate regressions | Single regression | Separate regressions | Single regression |
| | (1) | (2) | (3) | (4) |
| ***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 |
| 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 results of balance tests in the randomly assigned sample in Grade 7. The methodology is the same as that in Appendix Table A3.
*** p<0.01, ** p<0.05, * p<0.1.

Table A6: 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*** |

Table A6 (Continued)

| Teacher type | HRT | | SBT | |
|---|---|---|---|---|
| Specification | Separate regressions | Single regression | Separate regressions | Single regression |
| | (1) | (2) | (3) | (4) |
| ***Panel C. Dependent variable: Bachelor+*** | | | | |
| Age in 2013 | 0.52 | 0.58 | 0.01** | 0.00*** |
| 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 A3.
*** p<0.01, ** p<0.05, * p<0.1.

Table A7: Heterogeneity Analysis

| | Test scores | Cognitive ability | Noncognitive ability |
|---|---|---|---|
| | (1) | (2) | (3) |
| ***Panel A. By grade*** | | | |
| *Grade 7* | | | |
| HRT exper. | 0.11* | 0.14** | 0.09*** |
| | (0.06) | (0.06) | (0.03) |
| HRT female (=1) | 2.10** | 0.94 | 0.83 |
| | (0.89) | (1.05) | (0.50) |
| *Grade 8* | | | |
| HRT exper. | 0.12** | 0.16*** | 0.07* |
| | (0.05) | (0.06) | (0.04) |
| HRT female (=1) | 2.10** | 1.60* | 1.25*** |
| | (0.83) | (0.82) | (0.44) |
| ***Panel B. By student gender*** | | | |
| *Female* | | | |
| HRT exper. | 0.09* | 0.13** | 0.07* |
| | (0.05) | (0.05) | (0.04) |
| HRT female (=1) | 1.73** | 0.78 | 0.75 |
| | (0.75) | (0.81) | (0.49) |
| *Male* | | | |
| HRT exper. | 0.13** | 0.16*** | 0.08** |
| | (0.06) | (0.06) | (0.04) |
| HRT female (=1) | 2.37** | 1.67* | 1.30** |
| | (1.00) | (0.98) | (0.59) |
| ***Panel C. By type of residence (Urban v.s. Rural)*** | | | |
| *Urban* | | | |
| HRT exper. | 0.14** | 0.17** | 0.09** |
| | (0.06) | (0.06) | (0.04) |
| HRT female (=1) | 2.32** | 1.53 | 1.11** |
| | (1.03) | (1.09) | (0.51) |
| *Rural* | | | |
| HRT exper. | 0.08 | 0.10 | 0.02 |
| | (0.10) | (0.12) | (0.06) |
| HRT female (=1) | 1.61 | 0.57 | 0.66 |
| | (1.23) | (1.53) | (0.70) |

*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 A8: Teacher Characteristics and Student Performance (Additional Controls)

| | Test score | | Cognitive ability | | Noncognitive ability | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| HRT exper. | 0.11** | 0.10** | 0.14*** | 0.13*** | 0.07** | 0.07** |
| | (0.05) | (0.04) | (0.05) | (0.04) | (0.03) | (0.03) |
| HRT female (=1) | 2.05** | 2.02*** | 1.20 | 1.17 | 0.96** | 1.00** |
| | (0.79) | (0.73) | (0.81) | (0.76) | (0.40) | (0.39) |
| HRT bachelor+ (=1) | -0.65 | -0.65 | -0.89 | -0.88 | -0.37 | -0.41 |
| | (0.59) | (0.55) | (0.64) | (0.61) | (0.41) | (0.39) |
| HRT major educ. (=1) | -0.53 | -0.40 | -0.95 | -0.84 | 0.50 | 0.46 |
| | (0.80) | (0.79) | (0.71) | (0.80) | (0.72) | (0.72) |
| SBT exper. | | -0.00 | | -0.00 | | -0.05* |
| | | (0.05) | | (0.06) | | (0.03) |
| SBT female | | 0.18 | | 0.11 | | -0.41 |
| | | (1.01) | | (1.19) | | (0.70) |
| SBT bachelor+ | | -0.83 | | -0.68 | | -0.13 |
| | | (1.05) | | (1.11) | | (0.80) |
| SBT major educ. | | 3.44 | | 3.32 | | 0.43 |
| | | (2.33) | | (2.57) | | (1.44) |
| 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 A9: HRT Characteristics and Noncognitive Abilities (Big Five Measures)

| | Openness | | | | Conscientiousness | |
|---|---|---|---|---|---|---|
| | Hobbies | Curiosity | Fast learning | Fast reaction | Hardworking | Seldom escape |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| HRT exper. | 0.15 | 0.07 | 0.28** | 0.13 | 0.10 | 0.04 |
| | (0.10) | (0.07) | (0.11) | (0.08) | (0.13) | (0.03) |
| HRT female (=1) | 0.62 | -0.65 | 1.73 | 1.51 | 4.32** | 0.40 |
| | (1.28) | (0.68) | (1.54) | (1.07) | (1.81) | (0.52) |
| HRT bachelor+ (=1) | -0.73 | 0.51 | -1.65 | 0.74 | -2.34 | -0.73 |
| | (1.26) | (0.90) | (1.24) | (1.41) | (2.23) | (0.50) |
| HRT major educ. (=1) | 0.87 | -1.29 | 1.26 | 0.74 | 5.17** | -0.24 |
| | (1.46) | (0.99) | (2.66) | (2.91) | (2.19) | (0.76) |
| 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 | Agreeableness | Neuroticism | | | AES |
|---|---|---|---|---|---|---|
| | Social activities | Friendly & easygoing | Emotional stability | Confidence | Perception of Friendly classmates | Noncognitive abilities |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| HRT exper. | 0.41** | 0.09 | 0.03 | -0.00 | 0.15 | 0.00362** |
| | (0.17) | (0.09) | (0.10) | (0.10) | (0.11) | (0.00154) |
| HRT female (=1) | 5.56*** | 2.03 | 2.74** | 3.28*** | 0.52 | 0.0513** |
| | (1.91) | (1.38) | (1.26) | (1.12) | (1.40) | (0.0212) |
| HRT bachelor+ (=1) | -1.93 | -1.59 | 0.50 | -1.29 | 1.05 | -0.0186 |
| | (3.43) | (1.33) | (1.94) | (1.21) | (1.27) | (0.0204) |
| HRT major educ. (=1) | 0.87 | 0.70 | 3.21 | 1.09 | -0.09 | 0.0251 |
| | (1.67) | (2.84) | (2.81) | (1.85) | (2.01) | (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 17. 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 A10: HRT Characteristics and Student Test Scores (Reweighted)

| | Language | Math | Test score (reweighted) |
|---|---|---|---|
| | (1) | (2) | (3) |
| HRT exper. | 0.10* | 0.13** | 0.12** |
| | (0.05) | (0.05) | (0.05) |
| HRT female (=1) | 2.00** | 1.94** | 2.10** |
| | (0.84) | (0.77) | (0.83) |
| HRT bachelor+ (=1) | -0.82 | -0.04 | -0.44 |
| | (0.56) | (0.73) | (0.64) |
| HRT major educ. (=1) | -0.64 | -0.28 | -0.49 |
| | (0.78) | (0.84) | (0.82) |
| 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 70 and an SD of 10. 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 A11: HRT Characteristics (Grade 7) and Teacher-student Attrition (Grade 8)

| | HRT change (%) | Student attrition (%) | Student switch within a school (%) | Class reorganization (%) | Attrition in teacher-student pairs (%) |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| HRT exper. | 0.11 | 0.03 | -0.44 | -0.01 | -0.31 |
| | (0.48) | (0.07) | (0.37) | (0.01) | (0.49) |
| HRT female (=1) | 5.44 | 0.01 | -0.26 | -0.13 | 5.06 |
| | (5.72) | (0.68) | (3.57) | (0.14) | (5.11) |
| HRT bachelor+ (=1) | 0.01 | 0.05 | -1.99 | 0.05 | -1.87 |
| | (5.74) | (0.92) | (2.83) | (0.07) | (5.60) |
| HRT major educ. (=1) | -14.19 | 0.10 | 1.82 | 0.08 | -12.18 |
| | (9.13) | (1.26) | (4.02) | (0.08) | (10.02) |
| 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) missing from the survey, (2) switching from the original class within a school, (3) in reorganized classes, and (4) in classes with HRT changes. 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 A12: Robustness Excluding Classes with HRT Change and Using Current HRT Characteristics

| | Test score | Cognitive ability | Noncognitive ability |
|---|---|---|---|
| | (1) | (2) | (3) |
| ***Panel A. Excluding classes with HRT change*** | | | |
| HRT exper. | 0.11** | 0.14*** | 0.07** |
| | (0.05) | (0.05) | (0.03) |
| HRT female (=1) | 2.05** | 1.20 | 0.96** |
| | (0.79) | (0.81) | (0.40) |
| HRT bachelor+ (=1) | -0.65 | -0.89 | -0.37 |
| | (0.59) | (0.64) | (0.41) |
| HRT major educ. (=1) | -0.53 | -0.95 | 0.50 |
| | (0.80) | (0.71) | (0.72) |
| ***Panel B. Using current HRT characteristics*** | | | |
| HRT exper. | 0.13** | 0.14** | 0.08** |
| | (0.05) | (0.05) | (0.03) |
| HRT female (=1) | 1.94** | 1.24 | 0.93** |
| | (0.80) | (0.84) | (0.40) |
| HRT bachelor+ (=1) | -0.20 | -0.71 | -0.28 |
| | (0.52) | (0.64) | (0.43) |
| HRT major educ. (=1) | -0.74 | -0.70 | 0.85 |
| | (0.79) | (0.69) | (0.74) |
| | | | |
| 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 A13: Analysis of Variance

| Dependent variable: | Test score | Cognitive ability | Noncognitive ability |
|---|---|---|---|
| | (1) | (2) | (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 A14: Components and Factor Loadings of PCA in Mechanism Indices

| Component Analysis | | | Factor loadings of first component | |
|---|---|---|---|---|
| Component | Eigenvalue (1) | Proportion (2) | 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 |
| 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 A15: Summary Statistics of Mechanism Components

| | Observations | Mean | Standard deviation |
|---|---|---|---|
| ***Classroom environment (%)*** | | | |
| (1) Good atmosphere | 12,592 | 80.40 | 39.7 |
| (2) Friends study hard | 12,592 | 47.81 | 49.95 |
| (3) Friends study well | 12,592 | 44.37 | 49.68 |
| (4) Students seldom drink or smoke | 12,592 | 99.79 | 4.626 |
| (5) Students seldom go to Internet cafes | 12,592 | 95.23 | 21.32 |
| ***Motivation (%)*** | | | |
| (6) Bachelor degree or above | 12,232 | 83.80 | 36.84 |
| (7) Live in big cities | 12,232 | 91.20 | 28.34 |
| ***Parent-teacher interaction (%)*** | | | |
| (8) Teacher-parent communication | 12,286 | 8.32 | 27.62 |
| (9) Parent-teacher communication | 12,286 | 51.51 | 49.98 |
| (10) Harmonious relationship | 12,286 | 79.48 | 40.39 |
| (11) Parent-teacher meeting | 12,286 | 90.01 | 29.98 |
| (12) Parent fearless of communication | 12,286 | 76.40 | 42.47 |
| (13) HRT familiar with parents | 12,286 | 28.38 | 45.09 |
| (14) HRT respected by parents | 12,286 | 97.88 | 14.42 |
| ***Parental supervision (%)*** | | | |
| (15) Test score | 12,733 | 72.14 | 44.83 |
| (16) Dressing | 12,733 | 28.17 | 44.98 |
| (17) Watching TV | 12,733 | 44.63 | 49.71 |
| ***Parental expectation (%)*** | | | |
| (18) Bachelor degree or above | 12,232 | 83.80 | 36.84 |
| (19) Live in big cities | 12,232 | 91.20 | 28.34 |
| ***Teacher Effort*** | | | |
| (20) HRT working hours | 12,501 | 51.27 | 29.85 |
| (21) 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 A16: HRT Characteristics and Mechanism Indices (Alternative Measures)

| | Classroom environment | Student motivation | Parent-teacher interaction | Parental expectation | Parental supervision |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (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** |
| | (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 A17: HRT Characteristics and Student Performance (Inclusion of Mechanism Indices)

| | Test score | Cognitive ability | Noncognitive ability | Test score | Cognitive ability | Noncognitive ability |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| HRT exper. | 0.10** | 0.13** | 0.03 | 0.10** | 0.14*** | 0.03 |
| | (0.05) | (0.05) | (0.03) | (0.05) | (0.05) | (0.03) |
| HRT female (=1) | 1.87*** | 1.09 | 0.53 | 1.86*** | 1.10 | 0.54 |
| | (0.69) | (0.79) | (0.35) | (0.68) | (0.79) | (0.36) |
| Classroom environment | 0.10*** | 0.07*** | 0.19*** | 0.10*** | 0.07*** | 0.19*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Student motivation | 0.15*** | 0.09*** | 0.13*** | 0.22*** | 0.13*** | 0.17*** |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) |
| Parent-teacher interaction | 0.01 | 0.02** | 0.11*** | 0.01 | 0.03** | 0.11*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Parental supervision | 0.09*** | 0.03*** | 0.07*** | 0.09*** | 0.04*** | 0.07*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Parental expectation | 0.12*** | 0.08*** | 0.08*** | | | |
| | (0.01) | (0.01) | (0.01) | | | |
| 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 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 10. 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.