# TEACHER CHARACTERISTICS, TEACHER-STUDENT RELATIONSHIPS, AND STUDENT ACADEMIC OUTCOMES IN CHINESE JUNIOR HIGH SCHOOLS

by

CONGLI ZHANG

A DISSERTATION

Presented to the Department of Educational Methodology, Policy and Leadership
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Congli Zhang

Title: Teacher Characteristics, Teacher-Student Relationships, and Student Academic Outcomes in Chinese Junior High Schools

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Education Methodology, Policy, and Leadership by:

| | |
|---|---|
| Dr. David Liebowitz | Chairperson/Advisor |
| Dr. Kathleen Scalise | Core Member |
| Dr. Cengiz Zopluoglu | Core Member |
| Dr. Atika Khurana | Institutional Representative |

and

| | |
|---|---|
| Krista Chronister | Vice Provost for Graduate Studies |

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

DISSERTATION ABSTRACT

Congli Zhang

Doctor of Philosophy

Department of Educational Methodology, Policy and Leadership

June 2023

Title: Teacher Characteristics, Teacher-Student Relationships, and Student Academic Outcomes in Chinese Junior High Schools

Experimental evidence of the effects of teacher characteristics and teacher-student relationships on student performance is limited and even more scarce in education contexts outside of the United States. In this dissertation, I implement quasi-experimental research designs in two separate studies to investigate teacher-characteristic effects and teacher-student-relationship effects in the population of Chinese junior high school students. I draw analytic samples from a two-year, student-level, nationally representative dataset and leverage a national trend of random teacher-student assignments to investigate teacher effects on student performance as well as subject-specific self-concept. I estimate teacher effects as the within-school, between-teacher variance components of teachers' value added to student outcomes over a school year. In my first study, I find that, in China, more years of education or of teaching experience generally does not have a causal impact on student learning. Further, early career (less than three years) teachers consistently outperform their colleagues at the same school. Moreover, I detect some heterogeneity in teacher characteristic effects across subject areas: students benefit from teachers' graduate-level degree and Education major in Chinese (language arts) but learn less from math teachers who hold a graduate-level degree, with the effect sizes medium to large in magnitude. My second study first adds novel evidence about a national policy initiative in China: assigning a formal advisor role to a core-content teacher. I find that students

taught in their content area by their advisor had better relationships with their teacher, and students' self-concept in language subjects (Chinese, and English as the nationally mandated second language) and their math and English test scores were higher. In Chinese and English, the enhanced relationship between teachers and students caused by being taught by advisor consistently improved students' performance and the effect sizes were large in magnitude (although the estimates on Chinese score were imprecise). Together, these two articles contribute to the limited teacher effects literature in Chinese education context and importantly, provide implications for what teacher-level factors do or do not contribute to student performance to educators and policymakers worldwide.

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to Dr. David Liebowitz, thank you for your guidance, mentorship, and advocacy throughout my pursuit of this degree as well as the completion of this dissertation. I also would like to thank Dr. Kathleen Scalise, Dr. Cengiz Zopluoglu, and Dr. Atika Khurana for serving on my committee as well as your feedback and input throughout this work. Also sending my heartfelt appreciations to Dr. Ilana Umansky, Dr. Gina Biancarosa, and Dr. Joseph Stevens for your encouragement, friendship, and continuous support over my graduate school career. Together, to all the faculty members and graduate student friends – currently or previously at the University of Oregon – who have helped me grow as a researcher or as a person, thank you for bringing such precious memories to my life.

A special thank you to my previous colleagues and all the students I had taught back in the years when I was an educator in China. You have been the reason why I ask these research questions and the motivation behind my pursuit of educational research. I also extend my sincere respect and appreciations to the researchers and staff who contributed to the collection and documentation of China Education Panel Survey. Without the data you produced, the world will learn much less about the education system in China – the world's largest school system that serves over 250 million students.

All that I have accomplished would not have been possible without the unwavering love of my family. This space is too short to write down how much you have believed in me and unconditionally supported me in every step of my way. Thank you and I love you all. 谢谢你们的支持和鼓励，爸爸妈妈，妹妹，建强，Bill，柏宇，Ivy，和 Zoe。

To Boyu Liu.
Thank you for pushing through these hard years with me and
always saying “mum, I am proud of you”.

TABLE OF CONTENTS

## LIST OF FIGURES

LIST OF TABLES

# CHAPTER I. INTRODUCTION

Effectively hiring quality teachers and improving teacher performance are of central interest to education policy and practice, as it has become a bromide among observers of education policies that "teaching quality is the single most important school variable influencing student achievement" (McKenzie & Santiago, 2005, p.28). Indeed, over the past two decades, a large literature body has documented the outsize role teachers have in determining students' academic performance (Rockoff, 2004; Hanushek, 2011; Nye et al., 2004). Among this strand, one of the leading methods has been using teacher's value-added to student test scores as a proxy for teacher quality (Koedel et al., 2015). Using these approaches, a large body of literature identifies teacher quality as having consistent impact on students' math and reading achievement (Chetty et al., 2014; Kane et al., 2008; Rockoff, 2004) and later life outcomes (Jackson, 2018; Kraft, 2019). While establishing the link from the variation in teacher quality to meaningful changes in student outcomes is a helpful empirical fact, it does not identify specific teacher-relevant factors that drive teacher quality.

Understanding more about the causal effects of specific teacher-relevant factors on student performance has substantive implications to human resources policies and continuous improvement of schools and teachers. Responding to this call, a rich research stream seeks to understand whether strategic recruitment and selection policies could improve the average quality of incoming teachers by investigating whether observable teacher characteristics are systematically related to higher levels of student performance (e.g., Harris & Sass, 2011; Kane et al., 2008; Rockoff et al., 2011; Papay & Kraft, 2015). If certain teacher characteristics or credentials have consistent impacts on teacher productivity, schools could more effectively identify good teachers during recruitment and hiring, and teacher preparation and professional development programs could better prepare teachers and help them improve on their job.

Another line of research has focused on the social contexts of education and examined what school-based structures could provide more supportive teaching and learning environments, which in turn positively affect student learning. Here, the arguably most studied domain is teacher-student relationships as teachers (and the programs that prepare them) have long advocated for the importance of establishing strong relationships with their students as a mechanism to support student academic skill development. However, these important questions have been rarely examined in experimental ways, due to ethical and/or practical difficulties to implement randomized controlled trials in educational settings and the methodological challenges laid out by the prevalent sorting between students and teachers both within and across schools. Specifically, sorting of students to classrooms creates preexisting differences across classrooms, including unobserved differences related to student outcomes but impossible to be accounted for by simply adjusting for observable differences. Moreover, sorting of teachers to classrooms relates observed and unobserved teacher characteristics and skills to student outcomes, further clouding these preexisting differences. As a result, isolating the variation in student outcomes that is attributable only to a specific teacher factor has been a fundamental obstacle facing researchers attempting to make causal inference of their analysis results.

In my two-article dissertation, I build on these two lines of literature and overcome these methodological challenges to answer these research questions in careful quasi-experimental designs. Specifically, I leverage a natural experiment in China where junior high school (grades 7-9) students and teachers were randomly assigned to each other as a result of a nationwide education reform in 2006. This random assignment allows me to identify the exogenous variation in a teacher factor of interest, and then estimate the causal impacts of this factor without explicitly controlling for all other confounding variables that also affect student learning. It is worth noticing that my identification strategy takes advantage of the between-teacher comparisons, which is novel to the existing literature where the leading method to capture the impacts of teacher characteristics (such as years of experience) often relies on within-teacher variation to account for time-invariant factors outside teacher's control and improve internal validity (Koedel et al. 2015).

My first article examines which observed characteristics in a teacher's human capital profile (in particular, their level of post-secondary education and years of experience as a teacher) determine – or do not determine – teaching effectiveness. I find that teachers' total years of education does not affect student learning. However, Chinese (language arts) teacher's graduate-level degree and major in Education positively impacted students' performance and self-concept. In contrast, students of math teachers with graduate-level degree had lower performance. Moreover, I find that one extra year in teaching experience does not have a causal impact on student learning; in fact, the most effective teachers were early career teachers, which was particularly true for English (nationally mandated foreign language) and math. Last but not least, the finding of no marginal effect of teacher education or experience on student performance did not differ for students who are considered disadvantaged in Chinese society: students from low-income family or having a rural residency record (Zhao et al., 2017). Together, these findings suggest that although teacher human capital profile matters for Chinese language arts, teacher education and experience background in general may not provide reliable information in terms of identifying effective teachers; therefore, comprehensive human resources policies should look at more teacher attributes beyond these sorts of teacher characteristics.

The second article aims to address a long-standing research question: whether the observed associations between teacher-student relationships and student outcomes are causal, or solely due to self-selection of teachers and students. I started with documenting that, consistent with the existing literature, teacher-student relationship is substantially correlated with student learning across all core content subjects including Chinese, English, and math. However, these naïve estimates cannot be interpreted as causal due to omitted variable bias as well as reverse causality. To overcome these methodological barriers, I implement an instrumental variables approach to identify an exogenous portion of variation in teacher-student relationships that results from being taught by a randomly assigned teacher-advisor (a subject-matter teacher who also serves as the students' school-based advisor). I then use this exogenously determined teacher-student pairing to estimate the causal impacts of teacher-student relationships on student learning.

My first line of findings in this article is that being taught by a teacher-advisor has positive impacts on student learning, though with important nuances across different subject areas. It significantly improved student's performance and self-concept in English and relationship with English teachers; it also significantly improved student's self-concept in Chinese and relationship with Chinese teachers, but not test scores. In math, being assigned to a teacher-advisor increased students' test scores but had no impact on motivational or social-emotional outcomes, namely subject self-concept or relationship with teachers. Based on these effects of the exogenously decided teacher-advisor on teacher-student relationships in Chinese and English, my instrumental variables estimation identified large impacts of teacher-student relationships on student subject self-concept in both English and Chinese and small but also positive impacts on student performance in English. In math, unfortunately, since students' relationship with their teacher did not respond to being taught by advisor, whether teacher-student relationships matter for math performance remains unclear. In sum, these findings highlight the fact that social-emotional aspects of teaching brings meaningful change to students' learning in language subject areas and have substantive implications to policymakers and educators who seek evidence-based practices to improve student outcomes.

These two articles both add new, causal, between-teacher evidence to the teacher effects literature and more importantly, shed lights on what teacher-relevant factors drive – or equally importantly, do not drive – the causal impacts of content teachers on student learning. I present my dissertation in two stand-alone articles in Chapters II and III, before ending with a brief conclusion outlining my overall contributions as well as future research directions in Chapter IV.

## REFERENCES CITED


Chetty, Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. The American Economic Review, 104(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Hanushek. (2011). The economic value of higher teacher quality. Economics of Education Review, 30(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

Harris, & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. Journal of Public Economics, 95(7), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Jackson. (2018). What do test scores miss?: The importance of teacher effects on non-test score outcomes. Journal of Political Economy, 126(5), 2072–2107. https://doi.org/10.1086/699018

Kane, Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness?: Evidence from New York City. Economics of Education Review, 27(6), 615–631. https://doi.org/10.1016/j.econedurev.2007.05.005

Koedel, Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. Economics of Education Review, 47, 180–195. https://doi.org/10.1016/j.econedurev.2015.01.006

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. Journal of Human Resources, 54(1), 1-36.

McKenzie, & Santiago, P. (2005). Teachers Matter Attracting, Developing and Retaining Effective Teachers. OECD Publishing.

Nye, Konstantopoulos, S., & Hedges, L. V. (2004). How Large Are Teacher Effects? Educational Evaluation and Policy Analysis, 26(3), 237–257. https://doi.org/10.3102/01623737026003237

Papay, & Kraft, M. A. (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, 105–119. https://doi.org/10.1016/j.jpubeco.2015.02.008

Rockoff. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. The American Economic Review, 94(2), 247–252. https://doi.org/10.1257/0002828041302244

Rockoff, Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can You Recognize an Effective Teacher When You Recruit One? Education Finance and Policy, 6(1), 43–74. https://doi.org/10.1162/EDFP_a_00022

Zhao, Ye, J., Li, Z., & Xue, S. (2017). How and why do Chinese urban students outperform their rural counterparts? China Economic Review, 45, 103–123. https://doi.org/10.1016/j.chieco.2017.06.006

## CHAPTER II. THE CAUSAL IMPACTS OF TEACHER CHARACTERISTICS ON STUDENT ACADEMIC OUTCOMES

### Introduction

In most school human resources policies around the world, teacher human capital characteristics – such as education background and teaching experience – play an essential role. These characteristics, often referred to as "resume characteristics" or acquired characteristics (as opposed to sociodemographic characteristics such as gender and ethnicity), are widely used in policy initiatives such as hiring, pay, retention, and tenure decisions due to a number of unique properties they have: they are straightforward measures that are directly accessible in teacher's profile; they are practically and computationally cheaper than other teacher quality measures obtained from value-added models, classroom observations, and student/parent surveys; they can be easily and clearly categorized by certain policy thresholds, be directly influenced by education policy instruments, and they are relatively objective and less politically controversial. Among these characteristics-based policies, teacher compensation tied to experience is particularly common in many education systems. By offering higher salaries for experienced teachers, schools encourage teachers to retain and make the teaching profession more attractive to skillful candidates, in turn build a stable and experienced teaching staff that is beneficial to students as well as the whole school ecosystem.

Either explicitly or implicitly, these characteristics-based policies assume that teachers with stronger human capital profiles are better teachers, in other words, education attainment and experience improve a teacher's knowledge, skills, and productivity. This assumption might seem intuitive and straightforward; however, it is not fully supported by scientific research. In fact, an extensive empirical literature body suggests that human capital measures frequently used in teacher evaluation and compensation explain little of a teacher's contribution to student academic growth (Aaronson et al., 2007) and rarely any specific teacher characteristic has been identified as reliable predictor of student outcomes (Hanushek, 2011).

To date, the majority of the teacher characteristics literature is focused on Western (and particularly U.S.) education contexts. Moreover, teacher experience effect studies are largely led by teacher value-added method that relies on within-teacher variation to account for factors outside teacher's control. Whereas these approaches have incredibly improved internal validity, these within-teacher evidence is less informative to decision-making at the time of recruitment and hire, when decisions have to be made between candidates.

In this study, I turn to another education context – China, to fill in the eastern counterpart of our current state of knowledge. More importantly, I take advantage of a natural experiment in the nation and add more credibly causal, between-teacher evidence to the teacher characteristics effects literature. Specifically, I investigate the two guiding research questions in a quasi-experimental design: Whether—and to what extent—teacher education background affects student academic outcomes in China? Whether—and to what extent—years of teaching experience affects student academic outcomes in China?

## Literature Review

Whether and how to use teacher education background and teaching experience in human resources decisions are of great interests to policymakers, school administrators, and teachers themselves. These beg careful consideration of whether and to what extent various education and experience background characteristics contribute meaningfully to student learning. Unfortunately, researchers have so far not identified what teacher characteristics have consistent effects on student outcomes. For instance, one influential literature review of the education production function (Coleman, 1966) studies from last century, Hanushek (1986), failed to relate teacher preparation, experience, or salary with student achievement.

To gain a broad understanding of this topic from more recent literature, two systematic reviews, Wayne and Youngs (2003) and Coenen et al (2018) have synthesized researchers' findings at different time points and summarized some patterns emerging from the literature. Wayne and Youngs (2003) examined four categories of teacher characteristics: ratings of teachers' colleges, teachers' test scores, teachers' degrees and coursework, and teachers' certification status from 21 empirical studies in the US and highlighted that the teachers' college ratings and test scores seem to be positively related to student achievements. They also pointed out that findings on teachers' degree, coursework, and certification are inconclusive – with one exception: high school math teachers' certification, degree, and coursework are associated with increased student performance, which was confirmed by Coenen et al (2018). Note that Coenen et al (2018) extended their research scope to 58 studies selected from across the world. Their new insights included that the quality of teacher's college matters at secondary education level and teacher experience seems to contribute to student performance throughout a teacher's career.

In the U.S., over the past two decades, a growing body of experimental and quasi-experimental studies has brought additional information to our current state of knowledge. Consistently, findings on teacher education demonstrated a pattern that education attainment, major, and training programs generally do not add to student learning (Aaronson et al., 2007), with few exceptions. First, regarding advanced degree, math teachers who hold a graduate-level degree seem to positively impact student performance (Coenen et al., 2018; Guarino et al., 2013). Second, teacher's subject content knowledge proxied by licensure test scores was found to improve student learning (Clotfelter et al., 2006; Wayne & Youngs, 2003), particularly for math (Coenen et al., 2018).

The impacts of teacher experience on student outcomes are more nuanced. First, researchers generally agree that there is no consistent marginal effect (the effect of one additional year of experience on student learning throughout the distribution of teacher experience) of teacher experience on student outcomes (Aaronson et al., 2007; Hanushek, 2011; Rice, 2013). Instead, the relationships between teacher experience and student performance are evidently not linear (Clotfelter et al., 2006): whereas novice teachers are less effective than their more experienced counterparts (Harris and Sass, 2011), the impact of experience is in fact strongest in a teacher's early career (Papay and Kraft, 2015; Ladd & Sorensen, 2017). After the first few years, whether teachers continue to gain effectiveness is under debate: whereas some argue that the returns to experience level off (Kane et al., 2008; Rockoff, 2004), others find that teachers continue to improve throughout their full career (Harris & Sass, 2011), particularly in math (Papay & Kraft, 2015). Finally, there is also evidence that the relations between teacher experience and student learning differ across levels of education (Harris and Sass, 2011) and subject areas (Clotfelter et al., 2007; Rice, 2013; Coenen et al., 2018).

Empirical research conducted in Chinese education context is extremely scarce. In the very thin literature, researchers found that Chinese teacher's experience (Hu et al., 2022) and education background such as levels of education (Chu et al., 2015; Hu et al., 2022) do not add to student performance, which largely echoes the findings in the U.S. Beyond degree and experience, another commonly used teacher characteristic in China is teachers' professional rank in the local education agency's database, which also plays a practical role in schools' hiring, appointment, salary, and other human resources decisions. In order of prestige, teacher rank typically consists of four levels including novice (entry level), intermediate (level 2), advanced (level 1), and senior (high level). Each level has a relatively fixed number of teachers, which is determined by the local agency based on the full workforce in their database. Promoting teachers to higher levels is subject to local policy with various potential factors taken into decision metrics, such as degree, experience, performance, even publications and research experience, all making teacher rank an "ambiguous measure" (Chu et al., 2015). A more important note is that, since students' performance in high-stakes high-school entry exam at the end of junior high school serves as a critical measure of school quality and reputation in the community, schools oftentimes use rank promotion as an incentive to improve teacher performance. As a result, teacher rank is highly likely endogenous to student academic outcomes and without detailed information on local promotion policy, I am not able to unbiasedly estimate its effects in this study. Furthermore, to avoid introducing biases into my estimation, I do not include it as a teacher covariate either, which is different from some of the existing studies in this area.

In sum, the lack of empirical evidence in China, coupled with some debates around teacher characteristics effects, especially about the persistence of the returns to teacher experience, motivate the research questions in this article. Fortunately, the nationwide natural experiment makes it possible to obtain consistent estimates of these questions in a quasi-experimental design with stronger internal validity, which I discuss in further details in the following section.

## Background and School Settings

I conduct my research in China and identify my population of interest as Chinese public junior high school (grades 7-9) students and their core content teachers based on a critical policy consideration: Chinese public junior high schools are under a national law that drives the implementation of random assignment of teachers to students. To contextualize this natural experiment, it is helpful to note that China and many other countries such as France, Germany, India, Japan, and South Korea share a homeroom-based school system, which is different than the classroom-based settings in the U.S., UK, and many western countries. Specifically, unlike in the U.S. where each student has their own schedule and attends different classrooms each school day, Chinese students are grouped into homerooms, put on a shared homeroom schedule, and assigned a group of subject teachers who rotate to the homeroom to teach.

Throughout all years in which students attend the same school, students typically remain grouped with their original homeroom cohorts and their core content teachers, especially in subject areas that require three-year curriculum are encouraged to follow the homerooms rising to higher grades to gain familiarity of the full junior high curricula and teaching materials. This is particularly true for teachers who teach Chinese (language arts), English (nationally mandated foreign language), and math – the only three core subjects that not only require a full three-year education but also have largest weight over other subjects in the high-stakes high-school entry exam upon graduation. As a result, the common measure of experience is a three-year "teaching cycle" and teachers who finished their first teaching cycle are considered graduated from the apprentice stage.

**Natural Experiment Background**

In 2006, with great attention to education equality, the Compulsory Education Law (henceforth referred to as the 2006 Law; see Appendix B1 for more details) called off student tracking at all compulsory education levels (grades 1-9) and effectively eliminated national-, province-, and district-level academic exams below grade 9. The 2006 Law has stimulated a trend of random assignment of teachers to students across the nation, which was captured seven years later in the first nationally representative educational survey, the China Education Panel Survey (CEPS): 83% of the randomly sampled schools across the nation reported that they randomly assigned teachers to students upon students' entry to junior high school. Furthermore, researchers have documented this natural experiment in their studies investigating gender achievement gaps and teacher gender effects (Eble & Hu, 2020; Gong & Song, 2018; Xu & Li, 2018; ), peer effects (Xu et al., 2022), after-school tutoring (Sun et al., 2020), and equity issues in Chinese education (Zhao et al., 2017).

Both from the literature and my own observations as a formal school leader, the common teacher-student assignment approach has been that, supervised by local education departments, schools create either random or stratified homerooms of students upon students' entry to school, and then randomly assign teacher groups to homerooms (teachers are often assigned to multiple groups depending on their workload, for example, a math teacher is typically assigned to two homerooms because two classes per day, five days per week is the full time equivalent workload for a junior high school math teacher). Adding to the validity of the random assignment, local education departments typically review their public schools every year to check whether there are violations of the 2006 Law. Their strategies vary but many may require schools to submit a copy of their original homeroom rosters for the purpose of documentation. Others may conduct a student and/or parent survey or conduct more detailed school reviews in occasions when parents complain about unlawful student tracking or kids being discriminated against during homeroom assignment. These policy regulations greatly reinforce the validity of random assignment and in turn help it become an educational norm accepted by students, parents, and educators across the nation. This random assignment is crucial to my identification strategy and more evidence will be presented in further details in the Method section.

**Student Performance**

Teacher effect estimates will not be meaningful to policy and practice if the outcome variables are not valid measures of student performance. An issue at first sight is that, unlike in the US where student achievement tests are administered by the district or state, in China, none of the national-, province-, county-, and district-level tests below grade 9 exists, and each junior high school conducts its own locally developed tests to assess student performance. However, these test scores may, in fact, be valid measures of student learning due to two major reasons: (1) students are educated on the same grade-level knowledge and skills regardless their school and location since all compulsory education schools follow a national curriculum and most use the same PEP (People's Education Press) textbooks; and (2) school-administered tests are designed to be fair evaluations of teaching and learning progress because they are key assessments in a school's homeroom accountability system. Schools therefore often employ various approaches to achieve within-school validity and reliability, e.g., minimize test items not directly from current syllabus, avoid cheating or any types of manipulation of test score, include various types of items beyond multiple-choice questions to capture multiple dimensions of students' content knowledge and skills, to name just a few. The one remaining issue is that this high degree of within-school validity will be compromised across schools due to the large between-school variance: e.g., difficulty and quantity of test items are different (schools do not share test sheets) and scoring strategies vary (e.g., CEPS data indicates that schools were using a cap score of 100, 120, 130, or 150) from school to school. To address these issues, I standardized students' raw scores to have a mean of zero and standard deviation of one within each school and use only within-school variation in student score to estimate teacher effects.

Beyond exam score, I also include students' subject-specific self-concept, a measure that is rarely examined as an outcome variable in teacher effect literature. Self-concept is generally defined as "individuals' general perceptions of themselves in given domains of functioning" (Möller et al., 2009) and, from a social-cognitive perspective, is a critical variable in explaining student performance behavior (Marsh, 1986). I include it as an academic outcome variable based on two major considerations. First, with research showing the substantial correlations between student achievement and corresponding self-concept (Marsh et al., 2001; Möller et al., 2020), self-concept can serve as a robustness check to score outcome. More importantly, self-concept has its own research value in capturing the motivational dimensions of student learning as it feeds into performance, subject interest, educational decisions, and longer-term academic outcomes.

## Data and Measures

I draw my analytic sample from China Education Panel Survey (CEPS), China's first nationally representative, longitudinal survey of middle-school students and take advantage of its two waves of data. Starting in school year 2013-14, the CEPS team implemented a stratified, multi-stage sampling scheme to randomly select 112 junior high schools from across the country. Administrators from each randomly selected school were surveyed. Within each school, the sampling scheme then randomly selected two 7th grade and two 9th grade homerooms to survey. Within each homeroom, all students, parents, teacher-advisors, and content teachers in three core subjects (Chinese, English, and math) were surveyed. In school year 2014-15, most (n = 9,449, 91.93%) of the initial 7th grade cohort were successfully followed up in 8th grade, and these students will be the primary focus of my analysis. See Appendix A Data Description for more information about this data.

The two-wave CEPS data contains not only longitudinal information on a rich set of student-, family-, teacher-, and school-level variables but also whether the school randomly assigns teachers and students. Specifically, in the wave 1 survey, administrators were explicitly asked whether the school had randomly assigned teachers and students upon students' entry to middle school (before 7th grade began) and 83 percent (n=93) schools responded yes. This variable, coupled with the national random assignment trend stimulated by the 2006 Law, has been leveraged by researchers to overcome selection bias in estimating student outcomes under Chinese education context (e.g., Gong et al., 2018; Xu et al., 2022). I will further show detailed evidence of this random assignment in the Method section.

**Sample Restriction Process**

The validity of random teacher-student assignments is central to my identification strategy. However, CEPS data was not collected from a randomized controlled trial where researchers had full control of the teacher-student assignment, instead, the assignments fell under the purview of local school administrators and the data was self-report in nature. Acknowledging this data limitation, I implemented careful restriction criteria to obtain my analytic sample where students and teachers were the mostly likely to be truly randomly assigned to each other. Beforehand, I theorized three major contaminants of random assignment: (A) some non-public schools or under-resourced public schools still sorted students and teachers to meet specific groups' needs but reported random assignment on account of political incentives; in other schools who truthfully implemented random assignment, after assignment; (B) some parents lobbied their children to be placed in the homerooms with their desired teachers; and (C) under the pressure of homeroom accountability, some teachers removed (explicitly or through implicit encouragement) lower-achieving students from their homeroom (to other homerooms or another school) or schools used homeroom reassignment as some sort of policy intervention.

To deal with contaminants B and C, the most recent study using CEPS data (Xu et al., 2022) limited the data to only wave 1 information on the initial 7th graders in the 93 schools who reported random assignment, based on the rationale that 7th grade is the time when parents and teachers have the least knowledge about student academic ability therefore the least likely to sort students. This strategy may not be sufficient because CEPS wave 1 data was collected after the mid-semester test, i.e., 2-3 months after initial assignment, which leaves enough time for student sorting if the school indeed allowed it to happen. More importantly, CEPS' valuable asset, the two-wave longitudinal data, allows for the inclusion of prior scores in same and other subjects as the most important covariates to mitigate measurement error (Lockwood & McCaffrey, 2014) and reduce estimation bias (Chetty et al., 2014) – throwing it away is probably not a wise methodological decision.

I approached these three contaminants of random assignment in a different way and specify and justify my steps of sample restriction as follows. First, I limited sample schools to the 85 schools that were public schools (partially addressing contaminant A) and self-reported to have randomly assigned teachers to students before 7th grade began. I then moved on to address student sorting between wave 1 and wave 2. Note that more than 80 percent homerooms had at least some change in their membership between two waves but most of these changes were driven by students moving in or out of school, indicated by 830 (8.07%) students unable to follow up and 471 (4.75%) newcomers in wave 2 data (see Appendix A. Data Description). For identification purposes, I was relatively unconcerned about this across-school sorting because in all the models I fit, I would control for school fixed effects to absorb any time-invariant factors driving students to sort in or out of school. In contrast, I was concerned about within-school sorting (contaminants B and C), which will introduce considerable bias into the estimates of teacher effects. In dealing with this issue, I identified 22 schools that had at least one student change homeroom ID (but remain in the same school) between two waves and excluded them from my sample. I was left with 63 schools with two-wave data, which I used in my primary analyses throughout all three articles.

In Appendix C Table C1, I compare schools in my analytic sample (n=63) with the remaining schools (n=49) based upon observed descriptive statistics and show that these two groups of schools are indeed systematically different: my sample schools are more likely from coastal and urban area, serve a better educated population, and have smaller class sizes. This comparison suggests that excluding these 49 schools indeed helps address contaminant A. I believe sacrificing some degree of external validity in exchange for a much stronger internal validity is a sound decision and am more confident about the random assignment in my analytic sample. In the Method section, I will formally conduct a covariates balance check to provide empirical evaluation of random assignment validity based on wave 1 performance and student characteristics.

**Key Variables**

**Predictor Variables.** I model teacher education attainment and experience in different formats (continuous and categorical) to understand more beyond their linear relationship with student outcomes. Teacher education attainment is measured by three separate variables: education in years, an indicator for graduate degree, and an indicator for major in Education (i.e., pedagogy-centered majors such as Chinese Education, English Education, and Mathematics Education, as opposed to academic-focused majors such as Chinese Studies, English Studies, and

Mathematics Studies). Note that by the time of survey, the education background of the teacher workforce in the nation was a blend of associate degree in Education major, bachelor's degree in either Education major or academic-focused major, and graduate degree (highly likely academic-focused because Education major is 4-year capped). An examination of the highest (graduate) degree and Education major is helpful to shed light on different dimensions through which teacher education may affect students. The distributions of student observations by teacher education are displayed in Figure 1, showing that most teachers hold a bachelor's degree.

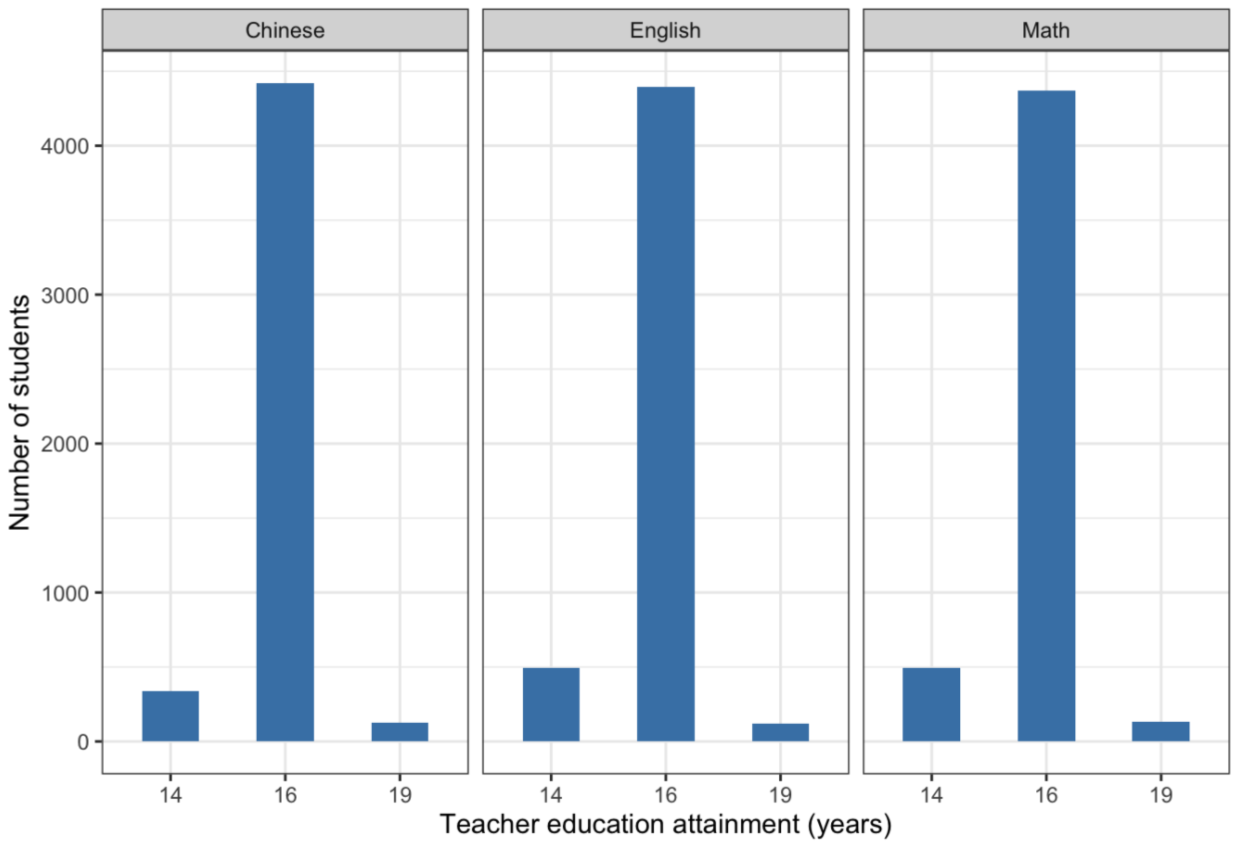



Figure 1. Distribution of student observations by teacher education attainment in years

*Notes*: Figure 1 presents the distribution of student observations by their teachers' education attainment measured in years, separately in each of the three subjects, Chinese, English, and math. The total years of education, 14, 16, 19 are equivalent to associate, bachelor, and graduate degree, respectively.

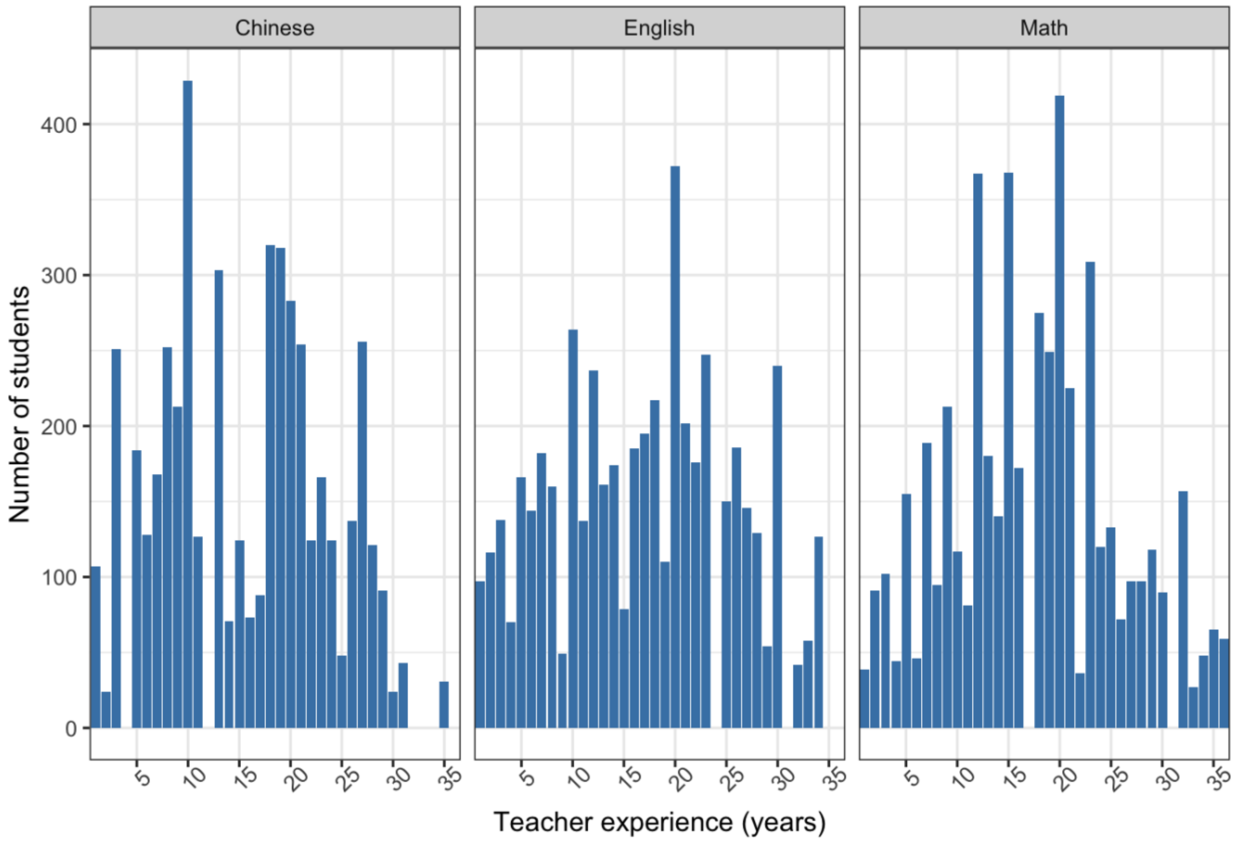


Figure 2. Distribution of student observations by teacher experience (in years)

*Notes*: Figure 2 presents the distribution of student observations by their teachers' experience measured in years, separately in each of the three subjects, Chinese, English, and math.

For teaching experience, my primary measure is a continuous variable measuring teacher's experience in years. As shown in Figure 2, the majority of the teachers in my sample have 15-25 years of experience. It is worth noticing that experience effects in my study are identified by between-teacher variations, which allows me to use experience measured in years to recover experience effects in the most intuitive and natural format – the change in student outcome corresponding to one additional year of experience. This improves upon traditional studies that rely on within-teacher variation therefore cannot directly use year as experience measure due to the perfect collinearity between year and experience.

To allow for nonlinearity in experience effects, I respecify the years of experience variable by collapsing years into three-year bins (first cycle 3-5 years, second cycle 6-8 years, and so forth, inspired by the three-year "teaching cycle" concept for Chinese core content teachers) to creates a series of dummy variables, with the left-out category being novice teacher who has two years or less experience. Ideally, this category should be no experience but because I am using within-school, between-teacher variation, there are unfortunately not enough observations in this category and the standard error will be problematically inflated. As a result, using each teaching cycle as a dummy variable allows me to estimate, compared to the "novice" teachers, how teachers with different numbers of teaching cycles under their belts impact students differently. It is important to point out that my assumption for using this strategy is that in the same school, teachers within the same teaching cycle do not differ meaningfully in terms of productivity. This is a different assumption than is made in traditional within-teacher studies of experience. Such studies deal with the collinearity between year and experience by collapsing years into a number of bins then using across-bin variation to estimate teacher experience effects (see Papay & Kraft, 2015 for a review of this method). These modeling approaches assume that compared to herself, a teacher's teaching effect does not change within each experience bin.

**Outcome Variables.** In each of the three subjects (Chinese, English, and math), student academic outcomes are measured by two variables, both of which contain unique information on student learning. First, due to the lack of national exams at junior high school level, I use students' subject-matter test score on school-administered mid-fall semester exam (obtained from their school records). The second outcome variable is subject-specific self-concept. The proxy available in the data is student's response to a 4-point Likert-scale survey item asking whether the subject is difficult. I reverse code the variable to represent four levels of self-concept: zero (very low), one (low), two (high), and three (very high). Overall, students report higher Chinese self-concept (70% reporting high or very high) than English and math selfconcepts (50 and 52% reporting high or very high). Note that both variables are standardized to have mean zero and unit variance within each school.

**Covariates.** I draw from wave 1 data three groups of covariates at the student-, homeroom-, and teacher-level to improve estimation precision. Student-level covariates include student wave 1 Chinese, English, math, and CEPS cognitive test scores as well as demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth (three categories including low-income, middle-class, and wealthy). The homeroom-level covariates include homeroom size and the homeroom means (leave-one-out mean, i.e., excluding self for each observation) of student characteristics. The teacher-level covariates include teacher gender and whether the teacher also serves as a teacher-advisor.

## Missing Data

Within each of the analytic sample schools (N = 63), I match students with core content teachers and obtain three separate samples for Chinese, English, and math and inspect the magnitude of missingness. On predictor variables, teacher education and experience, the missing rate is 0.92% and 4.78% for Chinese, 0% and 2.13% for English, and 0% and 2.63% for math. A closer look at Chinese sample shows that in school id number 92, only one Chinese teacher participated in the survey and also did not report any background information, which means that this school does not contribute to the estimates because I rely on within-school variation to recover teacher effects. This school is thus dropped from the Chinese sample. Across three subjects, the range of missing rate on outcome variables – test score and self-reported self-concept – is 1.08%-1.19% and 0.45%-0.54%. On student and teacher covariates, the missing rate is all below 2% except for student age, which is missing from 2.37% to 2.42%. Because of the relatively large sample size and small missingness, I assume the values are missing at random and drop all observations that have any missing value on my primary predictor and outcome variables, and I replace missing values on other variables with homeroom mean (for student covariates) or school mean (for teacher covariates).

The final sample size for Chinese, English, and math is 4,882, 5,010, and 4,995 students, respectively. I present summary statistics of key variables in Table 1.1. Note that the sample size and descriptive statistics of key variables are similar across three subjects with one exception, compared to Chinese and math teachers, English teachers were less likely to graduate from the Education major.

Table 1.1 Analytic sample summary statistics

| Key Variables | Chinese Sample | English Sample | Math Sample |
|---|---|---|---|
| | N = 4,882 | N = 5,010 | N = 4,995 |
| ***Predictor Variables*** | | | |
| Education (years) | 15.94 (0.71) | 15.87 (0.77) | 15.88 (0.79) |
| Graduate degree | 2.54% | 2.38% | 2.62% |
| Education major | 95.78% | 86.79% | 93.61% |
| Experience (years) | 16 (8) | 17 (9) | 17 (8) |
| ***Outcome Variables*** | | | |
| Score | 0.00 (0.99) | 0.01 (0.99) | 0.00 (0.99) |
| Confidence | 0.00 (0.99) | 0.00 (0.99) | 0.00 (0.99) |

*Notes*: Cells report mean and standard deviation for continuous variables and percentage of each category for categorical variables.

# Method

## Identification Strategy

To account for the possibility that higher-achieving students might sort to teachers with a stronger human capital profile, I leverage the random assignment of students and teachers not only enforced by the national regulation and reported by the surveyed schools (discussed in Introduction section) but also confirmed in the data. Specifically, I utilize the covariate balance check strategy frequently used in prior research (e.g., Clotfelter et al., 2006; Xu et al., 2022) and regress the measure of teacher educational attainment or years of experience against all student-level wave 1 covariates while controlling for school fixed-effects and clustering standard errors at the school level.

As demonstrated in Table 1.2, all the coefficients of my wave 1 covariates are small in magnitude and only a few of these tests are significant at conventional levels (likely due to sample idiosyncrasy). The small F-statistics also indicate that these covariates are jointly insignificant across students taught by teachers with different levels of education or experience. Importantly, none of the four wave 1 test scores seem to be significantly correlated with teacher education or experience. Thus, I conclude that the random assignment assumption required of my identification strategy is largely met. Provided random assignment, the variation in one of the teacher characteristics such as education or experience is independent from any observed and unobserved factors that also impact student outcomes; therefore, the estimated change in student outcomes can be attributed to the treatment – being taught a year by a teacher who has that certain characteristic.

To reenforce my identification strategy, I build on the teacher value-added literature and improve not only the precision but also the accuracy of my estimation by accounting for a set of the most important covariates – controls for prior achievement – in all the models I fit. The existing literature has informed me the good practice of how to choose from different measures of prior achievement and whether and how they contribute to the internal validity. Results from Chetty et al's (2014) quasi-experimental estimate of bias have shown that, the traditional model that only accounts for same-subject prior score may yield biased estimates but adding same- and other-subject scores from the prior year is a considerable improvement. In their context, adding more measures such as aggregates of prior achievement at classroom or school level improves little from adding same and other subject prior scores, so I did not choose this based on parsimonious consideration. In sum, I add the cubic function of prior year achievement in same and other subjects, meaning wave 1 Chinese, English, math, and CEPS cognitive test scores, in all the models I fit to capture varying functional forms of student prior learning ability as well as school and family inputs (Blazar & Kraft, 2017; Chetty et al., 2014; Kane et al., 2008; Kraft, 2019; Papay & Kraft, 2015).

Table 1.2. Covariates balance check: regressions of predictor variables on student wave 1 covariates

| | Education (years) | | | Experience (years) | | |
|---|---|---|---|---|---|---|
| | Chinese | English | Math | Chinese | English | Math |
| Wave 1 Chinese | 0.016 | -0.015 | -0.039 | 0.090 | -0.017 | -0.102 |
| | (0.025) | (0.023) | (0.023) | (0.226) | (0.186) | (0.239) |
| Wave 1 English | -0.019 | 0.054 | 0.052 | 0.249 | -0.090 | 0.117 |
| | (0.028) | (0.029) | (0.027) | (0.325) | (0.236) | (0.283) |
| Wave 1 math | 0.036 | -0.038 | 0.003 | -0.280 | 0.169 | -0.048 |
| | (0.021) | (0.027) | (0.023) | (0.290) | (0.208) | (0.249) |
| Wave 1 cognitive | -0.016 | 0.020 | -0.002 | -0.025 | 0.270 | -0.074 |
| | (0.021) | (0.028) | (0.015) | (0.306) | (0.209) | (0.296) |
| Female student | -0.015 | -0.017 | -0.017 | 0.081 | 0.246 | 0.099 |
| | (0.013) | (0.015) | (0.014) | (0.137) | (0.125) | (0.173) |
| Age | 0.013 | -0.011 | 0.016 | -0.480 | 0.094 | -0.140 |
| | (0.013) | (0.025) | (0.018) | (0.286) | (0.151) | (0.231) |
| Only child | -0.013 | -0.004 | -0.025 | 0.048 | -0.302 | 0.171 |
| | (0.019) | (0.012) | (0.014) | (0.181) | (0.169) | (0.211) |
| Rural residency | 0.016 | 0.006 | -0.031 | 0.292* | -0.266 | -0.091 |
| | (0.021) | (0.013) | (0.019) | (0.145) | (0.179) | (0.227) |
| Migrant family | -0.016 | 0.019 | 0.024 | 0.340 | -0.128 | -0.229 |
| | (0.027) | (0.021) | (0.018) | (0.184) | (0.191) | (0.268) |
| Mother education | 0.002 | -0.003 | -0.002 | -0.011 | -0.019 | 0.042 |
| | (0.004) | (0.003) | (0.003) | (0.032) | (0.028) | (0.042) |
| Father education | 0.004 | -0.002 | -0.001 | 0.061 | 0.009 | -0.014 |
| | (0.003) | (0.002) | (0.002) | (0.033) | (0.023) | (0.033) |
| Family income | -0.006 | 0.011 | 0.059* | 0.022 | 0.153 | -0.518* |
| | (0.021) | (0.022) | (0.023) | (0.272) | (0.149) | (0.234) |
| School FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| F-Statistics | 0.853 (df = 12; 61) | 1.010 (df = 12; 62) | 1.141 (df = 12; 62) | 1.731 (df = 12; 61) | 1.791 (df = 12; 62) | 1.347 (df = 12; 62) |
| Observations | 4882 | 5010 | 4995 | 4882 | 5010 | 4995 |
| R2 | 0.616 | 0.651 | 0.623 | 0.551 | 0.762 | 0.539 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where one of the predictor variables (teacher education and experience measured in years) is regressed on student score measures and characteristics from wave 1 data. All models control for school fixed effects and cluster standard errors at school level.

Based on this identification strategy, I recover the effect of teacher education or experience by estimating a linear value-added model in the following using ordinary least squares (OLS) regression:

$$A_{it} = \alpha_g (g (A_{i,t-1})) + \beta CHAR_{jt} + \gamma X_{i,t-1} + \delta P_{j,t-1} + \lambda T_{j,t-1} + \theta_s + \varepsilon_i$$

where i, j, s, t denote student, teacher (homeroom), school, year; $A_{it}$ is student i's academic performance or self-concept in year t; $CHAR_{jt}$ is teacher j's education or experience variable; g ($A_{i,t-1}$) is the cubic functions of student i' s prior score in Chinese, English, math, and CEPS cognitive test; $X_{i,t-1}$, $P_{j,t-1}$, and $T_{j,t-1}$ are previously discussed wave 1 student-, homeroom peer-level, and teacher-level covariates in year t-1; and $\varepsilon_i$ is the idiosyncratic error term. The coefficient of interest is β, which is the estimated effect of a given teacher education or experience variable on a given student outcome. Note that I control for school fixed effects ($\theta_s$) to account for school time-invariant characteristics that include both students' and teachers' sorting to schools, then cluster standard errors at the school level to account for the within-school correlations among residuals. I estimate each of the three subjects, Chinese, English, and math, separately.

## Results

### Effects of Teacher Education on Student Learning

In Table 1.3, I report the estimates of differences in student subject score as well as self-concept that corresponds to one additional year in teacher education attainment. In all model specifications, I adjust for the cubic function of student wave 1 score in Chinese, English, math, and CEPS cognitive test as well as school-fixed effects, and cluster standard errors at school level. To test the robustness of my specifications, I estimated each subject three times, as shown in three columns; I adjust for student's own demographic characteristics in column 1, then add aggregates of these characteristics at student's homeroom peer level in column 2, and lastly, further add teacher covariates in column 3. The third specification is preferred. The results show that one additional year in teacher's education attainment (range from 14-19 years, see Figure 1) generally did not have causal impact on student learning across three subjects with a couple of exceptions: it curbed student self-concept in math by 0.140 standard deviation (SD) and in contrast, it meaningfully improved Chinese score but the estimates were statistically imprecise.

Additionally, as shown in Appendix Table C2, there was no heterogeneous effect for student groups that are considered disadvantaged in Chinese society: students from low-income families and students who hold a rural residency status (Zhao et al., 2017).

I further examined whether and to what extent teacher's graduate-level degree or majoring in Education affect student learning and show my results in Tables 1.4 and 1.5. All model specifications are similar to those in Table 1.3. Related to the findings in Table 1.3, math teacher's graduate level education decreased student score by 0.202 SD. A new finding is that for Chinese teachers, those having a graduate degree significantly improved student self-concept by 0.359 SD and those graduating from educational studies major had marginal but positive effects on student score and self-concept.

**Effects of Teacher Experience on Student Learning**

Table 1.6 demonstrates the estimates of the effects of teacher experience on student subject score and self-concept in Chinese, English, and math. Consistently across different model specifications, one additional year in teaching did not have significant impact on student learning. . Additionally, as shown in Appendix Table C3, there was not heterogeneous effect for students from low-income families and students who hold a rural residency status.

A further investigation of whether teachers having finished different amount of teaching cycles (3-5 years, 6-8 years, 9-11 years, and so forth) are more effective than their novice school colleagues, I respecify experience as a set of dummy variables that each indicates a given cycle and visually display the point estimate and corresponding 95 percent confidence intervals for each cycle in Figures 3 and 4. Detailed point estimates and corresponding standard errors can be found in Appendix Table C4. The overall pattern is that more experienced teachers were in fact less effective than novice teachers who were still working on their first cycle. This pattern was more persistent for English than for the other two subjects.

## Discussions and Policy Implications

Schools all over the world face the challenge of hiring effective teachers, helping teachers grow, and making high-stakes personnel decisions such as tenure, ranking, and turnover. It is intuitive and convenient for governments, education agencies, and individuals to invest in education, preparation, and experience when it comes to identifying and developing effective teachers. However, consistent with the Western literature, my findings from Chinese junior high school teachers do not support the intuition that more years of education (beyond 14 years/associate degree) or experience in teaching contribute to student's performance or self-concept in all core content subjects.

Table 1.3. Effects of teacher education (in years) on student academic outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Education (yrs) | 0.053 | 0.042 | 0.040 | 0.027 | 0.019 | 0.025 | -0.026 | -0.020 | -0.021 |
| | (0.045) | (0.050) | (0.065) | (0.027) | (0.025) | (0.020) | (0.026) | (0.032) | (0.032) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.598 | 0.601 | 0.601 | 0.701 | 0.702 | 0.703 | 0.608 | 0.609 | 0.612 |

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Education (yrs) | 0.019 | 0.010 | 0.016 | 0.004 | -0.016 | -0.007 | -0.071* | -0.134*** | -0.140*** |
| | (0.072) | (0.063) | (0.049) | (0.028) | (0.029) | (0.030) | (0.033) | (0.036) | (0.037) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.075 | 0.079 | 0.083 | 0.253 | 0.256 | 0.260 | 0.238 | 0.244 | 0.244 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where subject score or self-concept is regressed against teacher education measured in years. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

Table 1.4. Effects of teacher graduate degree on student academic outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Graduate degree | 0.230 | 0.239 | 0.238 | 0.027 | 0.000 | 0.040 | -0.219* | -0.239 | -0.202* |
| | (0.210) | (0.191) | (0.216) | (0.060) | (0.065) | (0.049) | (0.091) | (0.120) | (0.083) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.599 | 0.601 | 0.602 | 0.701 | 0.702 | 0.703 | 0.608 | 0.610 | 0.612 |

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Graduate degree | 0.302 | 0.313 | 0.359* | 0.067 | -0.013 | 0.055 | -0.026 | -0.150 | -0.135 |
| | (0.169) | (0.163) | (0.149) | (0.108) | (0.143) | (0.142) | (0.094) | (0.169) | (0.177) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.076 | 0.080 | 0.085 | 0.253 | 0.256 | 0.260 | 0.237 | 0.241 | 0.241 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where subject score or self-concept is regressed against the indicator of teachers' graduate-level degree. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

Table 1.5. Effects of teachers' Education major on student academic outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Education major | 0.312** | 0.272 | 0.244 | -0.007 | -0.039 | -0.034 | 0.012 | -0.004 | 0.046 |
| | (0.104) | (0.149) | (0.157) | (0.070) | (0.071) | (0.050) | (0.106) | (0.101) | (0.072) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.599 | 0.602 | 0.602 | 0.701 | 0.702 | 0.703 | 0.608 | 0.609 | 0.612 |

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Education major | 0.414*** | 0.412* | 0.278 | -0.011 | -0.043 | -0.034 | -0.109 | -0.100 | -0.083 |
| | (0.115) | (0.164) | (0.193) | (0.087) | (0.091) | (0.048) | (0.136) | (0.111) | (0.108) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.077 | 0.081 | 0.084 | 0.253 | 0.256 | 0.260 | 0.237 | 0.241 | 0.241 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where subject score or self-concept is regressed against the indicator of teacher majoring in Education. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

Table 1.6. Effects of teacher experience on student academic outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Experience (yrs) | -0.002 | 0.000 | 0.002 | -0.003 | -0.002 | -0.004 | -0.004 | -0.005 | -0.004 |
| | (0.004) | (0.004) | (0.004) | (0.003) | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.598 | 0.601 | 0.601 | 0.701 | 0.702 | 0.703 | 0.608 | 0.610 | 0.612 |

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Experience (yrs) | -0.005 | -0.002 | 0.000 | 0.003 | 0.002 | 0.000 | -0.004 | -0.003 | -0.002 |
| | (0.006) | (0.005) | (0.005) | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.003) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.076 | 0.079 | 0.083 | 0.253 | 0.256 | 0.260 | 0.237 | 0.241 | 0.241 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where subject score or self-concept is regressed against teacher experience measured in years. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

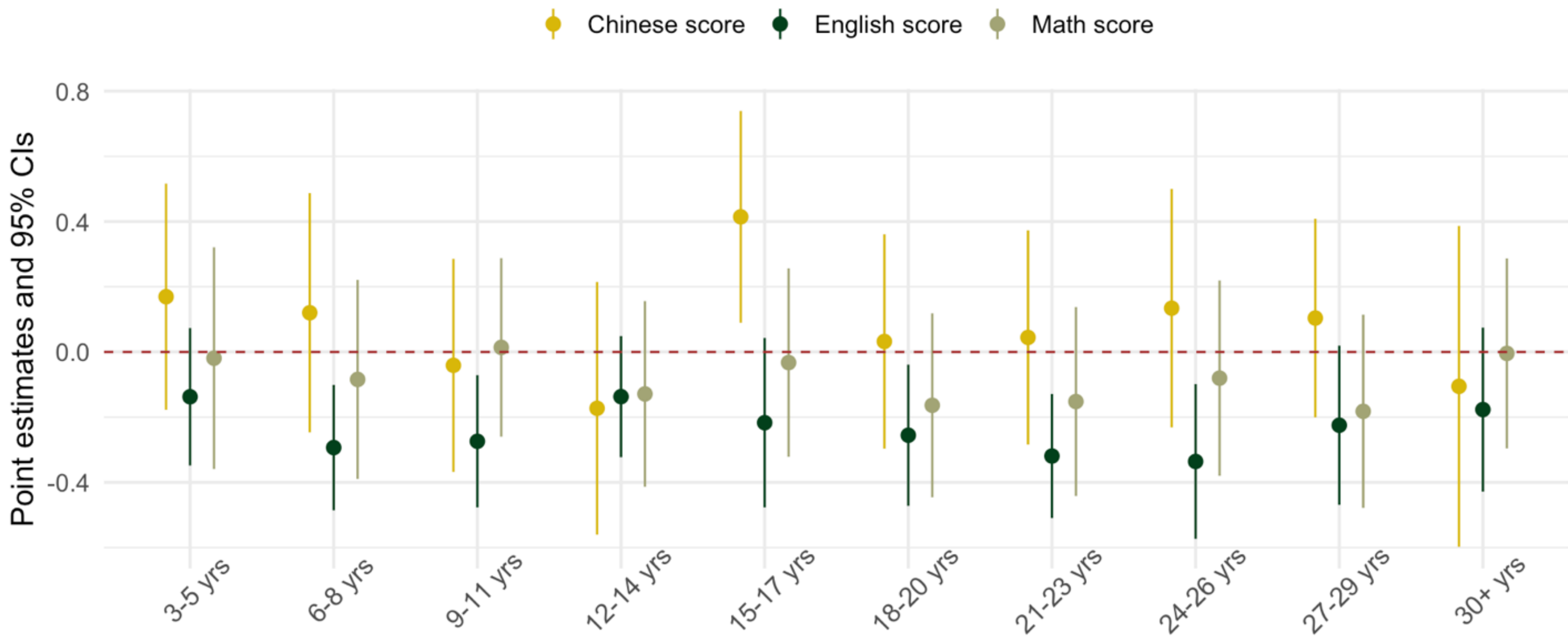



Figure 3. Nonlinear effects of teaching experience on subject score

*Notes*: Figure 3 demonstrates the average relationship between the listed bins of teacher experience and students' Chinese, English, and math score outcomes from an ordinary least squares regression estimator, which regresses each score outcome on a set of indicators that each represents a year bin (0-2 years as reference level), after accounting for student-, homeroom-, and teacher-level covariates as well as school fixed-effects. Standard errors are clustered at school level.

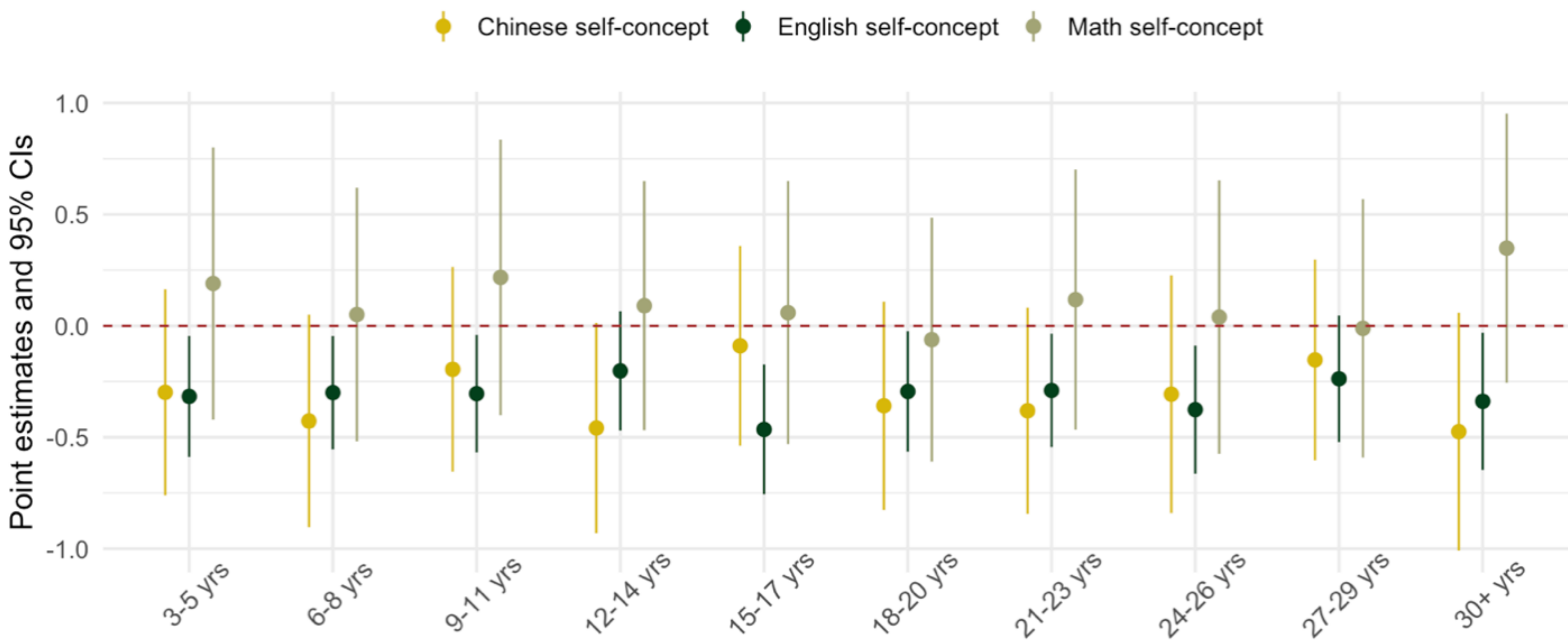



Figure 4. Nonlinear effects of teaching experience on self-concept

*Notes*: Figure 4 demonstrates the average relationship between the listed bins of teacher experience and students' self-concept outcomes in Chinese, English, and math from an ordinary least squares regression estimator, which regresses each outcome on a set of indicators that each represents a year bin (0-2 years as reference level), after accounting for student-, homeroom-, and teacher-level covariates as well as school fixed-effects. Standard errors are clustered at school level.

The impacts of teacher education or experience background on student learning are complicated and nuanced as they should be, because they are confounded by numerous observed and unobserved factors – from every aspect of student's learning and living environment – that also influence student learning. In this study, I leverage the random teacher-student assignment and between-teacher comparison condition in each school in a nationally representative, longitudinal dataset of Chinese junior high schools and recover the effects of teacher education and experience characteristics on student academic outcomes.

Note that this quasi-experimental design effectively accounts for the sorting of individual students to teachers but does not rule out the possibility that teachers can influence individual students through peer effects (e.g., a student learned more with a skillful teacher not only because her teacher was effective in teaching but also her homeroom cohorts were all making progress with this teacher), though this issue is mitigated by the various homeroom-level covariates I include in my estimation. Nonetheless, it is helpful to bear in mind that the teacher effect in my study is broadly defined – it is blend of teacher's direct effect and indirect effect (through homeroom peers) on individual students. Another note is that the estimates of teacher experience effects do not fully account for teacher-cohort impact, i.e., the influence of hiring policy or (more or less) preparation in certain year(s) for a certain cohort of teachers, although this issue is less concerning because of the relatively stable policy environment and similarity in teachers' background within each school.

My study has three major contributions to the empirical literature of teacher effects. First, the effects of teacher education and preparation on student learning outcomes differ considerably across subject areas. Chinese teachers' graduate-level degree and major in Education consistently improved student self-concept and students scored higher if their teachers graduated from education studies major. Math teachers with a graduate-level degree, on the contrary, negatively impacted student performance and more years of education decreased student self-concept. For English teachers, years of education, degree level, and major did not seem to matter.

It is worth noticing that this overall pattern of negative effects of teacher education background in math is different from the findings in the US context where researchers generally found greater variability in teacher effects in math than in reading (Hanushek & Rivkin, 2010; Papay & Kraft, 2015). This is not surprising because a growing line of literature has started to document some fundamental differences between the education systems of developing and developed countries, for instance, the weaker associations between family socioeconomic status and achievement in developing compared to developed countries (Kim et al., 2019) and the even weaker effect sizes for math/science than for language (Chinese/English) subjects in China (Liu et al., 2020). My finding further indicates that the mechanisms through which math teachers affect student learning may be different between China and the US. For instance, an anecdotal knowledge of math education in China is that students can improve their math score by repeated practicing (e.g., students often score higher by accurately perform complex hand calculations since the nation does not allow any calculator usage in high-stakes high-school and college entry exams). One consequence is that some teachers exploit this knowledge and improve their students' performance by assigning large amount of homework and overpreparing students for exams. Another consequence is that some parents, pressured by the highly competitive education system, seek afterschool tutoring or individual attention from teachers for their kids, which is confirmed in the CEPS data where among all students in the nation, 25.8% reported that they attended afterschool tutoring in math, whereas the proportions were 10.6% and 22.5% in Chinese and English.

My second major finding is that, there is no causal effect of years of teaching on student learning outcomes, neither on score nor self-concept, across all three subjects. An exploration of nonlinearity in experience effects showed that more experienced English teachers consistently underperformed their novice colleagues whose experience was less than two years. This pattern was less obvious for math and Chinese teachers, but it was evident that across all subjects, more experienced teachers did not have better student outcomes in comparison to their novice colleagues. This nonlinear effect differs from the recent findings in the US context where gains from experience were found to be largest in a teacher's first few years then persist throughout the teacher's full career (Harris & Sass, 2011; Papay & Kraft, 2015) but lines up with a few earlier studies that found no returns to experience (Aaronson et al., 2007) or no returns to experience in a teacher's later career (Boyd et al., 2008; Kane et al., 2008).

The biggest concern regarding this counterintuitive finding is that it might be a result of more experienced teachers systematically leaving for better careers. Although the data limitation restrains me to fully investigate this matter, I have two arguments to strengthen my estimation. First, the inclusion of school fixed effects in my estimator effectively accounts for time-invariant school factors that drive the systematic sorting of teachers in or out of school. Moreover, the data did show that among the 63 schools in my analytic sample, eight (13 percent) reported a relatively high rate of teacher turnover. I excluded these schools and show in Appendix C Figures 5 and 6 that this finding held for schools without high rates of teacher turnover.

Third, I did not find heterogeneous effects of teacher education or experience on student learning. Specifically, the fact that teacher education attainment and experience do not have marginal effect on student performance persists for student groups that are considered disadvantaged in Chinese society: students from a low-income family and students who have a rural residency record.

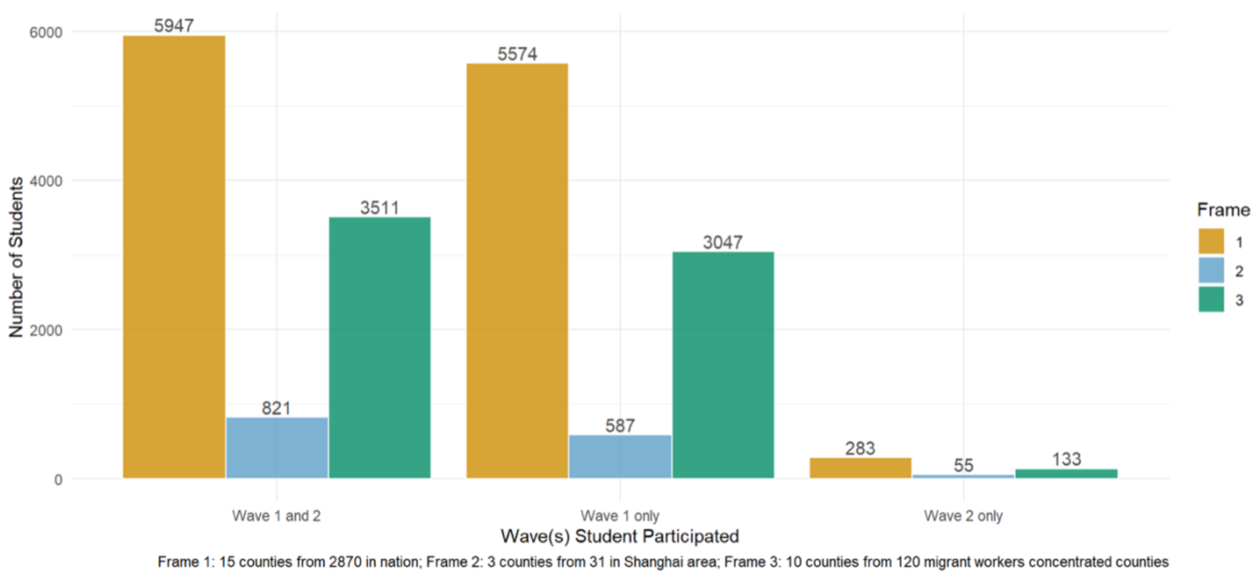
These findings should be interpreted with considerations of an important education context. Specifically, the effects of teacher education and experience background could be pulled toward null direction and similarly, the effects of new teachers could be boosted by the frequent and in-depth collaborations and peer-learning among teachers in China. From both observations of school practice and review of literature, subject-based professional learning communities (PLCs) among teachers – such as mentoring and coaching relationships, classroom observation and feedback workshops, and regular meetings on curriculum/material development, class preparation, and knowledge sharing – are highly valued, common practice in Chinese schools (Chen, 2020; Liu & Hallinger, 2018). This unique professional environment was confirmed in CEPS survey, where 96.2 percent (N = 761) of the teachers in the nation indicated that they had frequent communications with colleagues on teaching-related matters including curriculum and materials (69 percent of the 761 teachers), teaching methods (86.3 percent), developing quizzes and exam sheets (32.6 percent), and classroom management and student development (80.2 percent). As a result, it is possible that teachers who are novice or from less prepared background have the sufficient professional and structural support from their PLCs therefore are able to quickly catch up with or even outperform their experienced or skillful colleagues.

My study suggests that education decisionmakers should strategically spend resources on recruitment and hiring and be mindful that 1) teachers with stronger human capital profile are not necessarily better teachers, 2) teacher education background matters more for Chinese than other subject teachers, and 3) teaching experience should not be a policy focus. In fact, in terms of a more evidence-based hiring metric, it may worth including more characteristics and attributes of teachers such as content knowledge for teaching, cognitive skills, and noncognitive skills. For example, Rockoff et al (2011) studied a variety of nontraditional teacher measures including teaching-specific content knowledge, cognitive ability, personality traits, feelings of self-efficacy, and scores on a commercial teacher-screening instrument – in addition to traditional characteristics collected by local agency. They found that adding nontraditional measures to traditional measures indeed explains more variation in predicting teacher effectiveness and both researcher-measured teacher cognitive and noncognitive skills were both significant predictors of student achievement. Moreover, teacher demographic characteristics that match with current students can also add to school outcomes as a growing literature body has identified that teacher-student gender match (Dee, 2005; Egalite & Kisida, 2018; Xu & Li, 2018) and racial match (Dee, 2004; Egalite & Kisida, 2018).

## APPENDIX A. DATA DESCRIPTION

Conducted by the National Survey Research Central (NSRC) of Renmin University, China, the China Education Panel Survey (CEPS) started in school year 2013-2014 and employed a stratified, four-step random sampling procedure to draw a random sample of middle schools, teachers, and students from the nation. First, they randomly selected 28 school districts/counties with probability proportional to size (PPS) from three stratified sample frames, specifically, 15 from 2,870 districts/counties (frame 1) in the nation, 3 from 31 districts/counties in Shanghai area (frame 2), and 10 from 120 migrant labor concentrated districts/counties (frame 3). Second, within each district/county, they randomly selected four schools from all schools serving 7th and/or 9th grades with PPS. Third, within each school, they randomly selected two homerooms from 7th grade and another two from 9th grade. Fourth, within each homeroom, they included all students and administered separate surveys to students, parents, homeroom advisory teachers, classroom teachers for three core subjects (math, Chinese, and English), and school administrators.

Using this procedure, the CEPS team surveyed 112 schools with their 10,279 7th grade and 9,568 9th grade students in school year 2013-14 and successfully followed up with 9,449 of the original 7th graders (follow-up rate 91.9%) along with 471 new students in school year 2014-15. Detailed numbers of students by wave and frame are visualized in the following bar chat. Note that the 9,449 students with two-wave data (the first three bars) were the focus of my dissertation, see more discussion in the text.



# APPENDIX B. EDUCATION POLICIES

## B1. Compulsory Education Law (2006)

The Compulsory Education Law[1] was amended and adopted at the 22nd Session of the 10th National People's Congress Standing Committee and issued as No. 52 Order of the President on June 29, 2006. Relevant to my research, the law highlighted that all school-age children and adolescents shall have equal right and the obligation to receive a 9-year compulsory education (Article 4) at the schools near their residency (Article 12). They shall go to school without taking any examination (Article 12). The county level governments and education departments shall promote the balanced development among schools and narrow down school quality gaps (Article 22). No education government may create key schools and non-key schools and no school may create key classes and non-key classes (Article 22). No school may expel students based on school management rules (Article 27). Legal liabilities are attached to the violations of these articles.

## B2. Regulations of Advisory Teachers by Ministry of Education (2009)

The Ministry of Education issued the Regulations of Advisory Teachers[2] on August 12, 2009. Relevant to my research, the regulation specified advisory teacher's core responsibilities as moral education, student discipline, student development, and mentoring. The regulation emphasized that every homeroom in the country shall have an advisory teacher and the position is half-time equivalent. A homeroom's advisory teacher should teach the homeroom and should be ethical, psychologically healthy, caring, dedicated, and having strong communication ability and managerial skills.

1. See http://www.lawinfochina.com/Display.aspx?lib=law&Cgid=77520 for a translation of the Law.
2. No translation of this document was found on the internet. The Chinese version is here http://www.moe.gov.cn/srcsite/A06/s3325/200908/t20090812_81878.html.

APPENDIX C. ADDITIONAL TABLES AND FIGURES

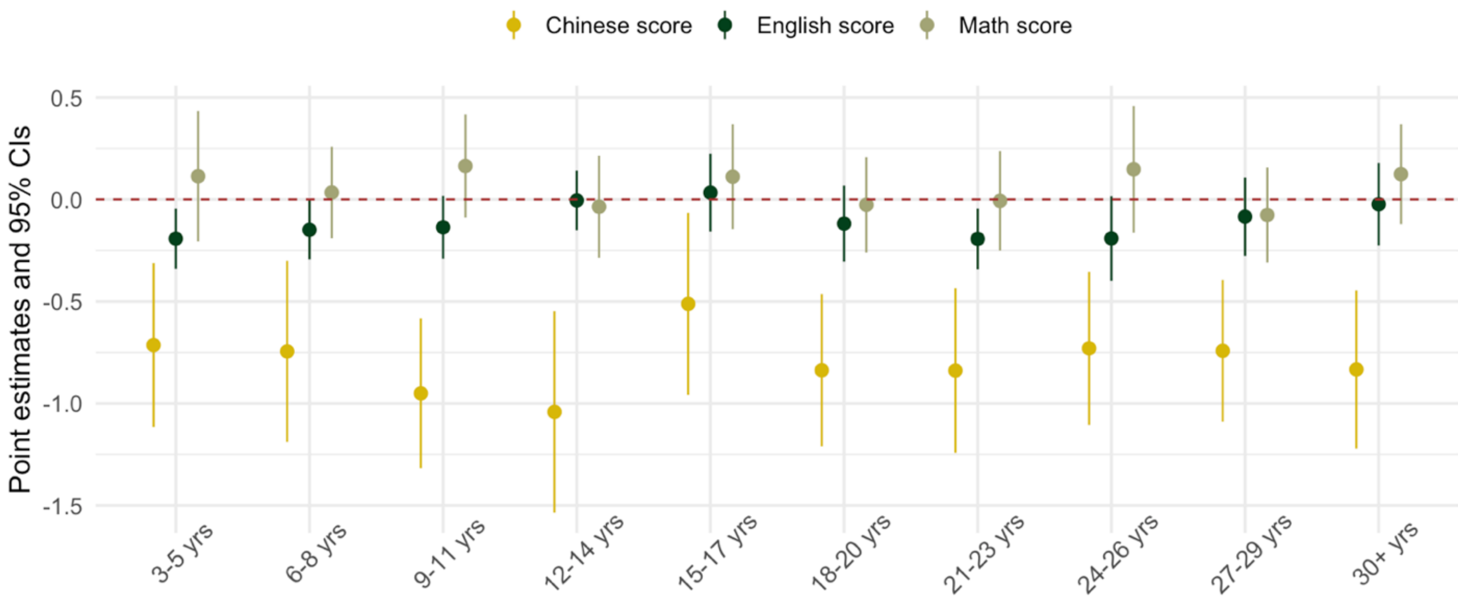



Figure 5. Experience on score for schools without high turnover

*Notes*: Figure 5 demonstrates the average relationship between the listed bins of teacher experience and students' score outcomes in Chinese, English, and math from an ordinary least squares regression estimator, which regresses each outcome on a set of indicators that each represents a year bin (0-2 years as reference level), after accounting for student-, homeroom-, and teacher-level covariates as well as school fixed-effects. Standard errors are clustered at school level. Note that the schools reporting high levels of teacher turnover are excluded.

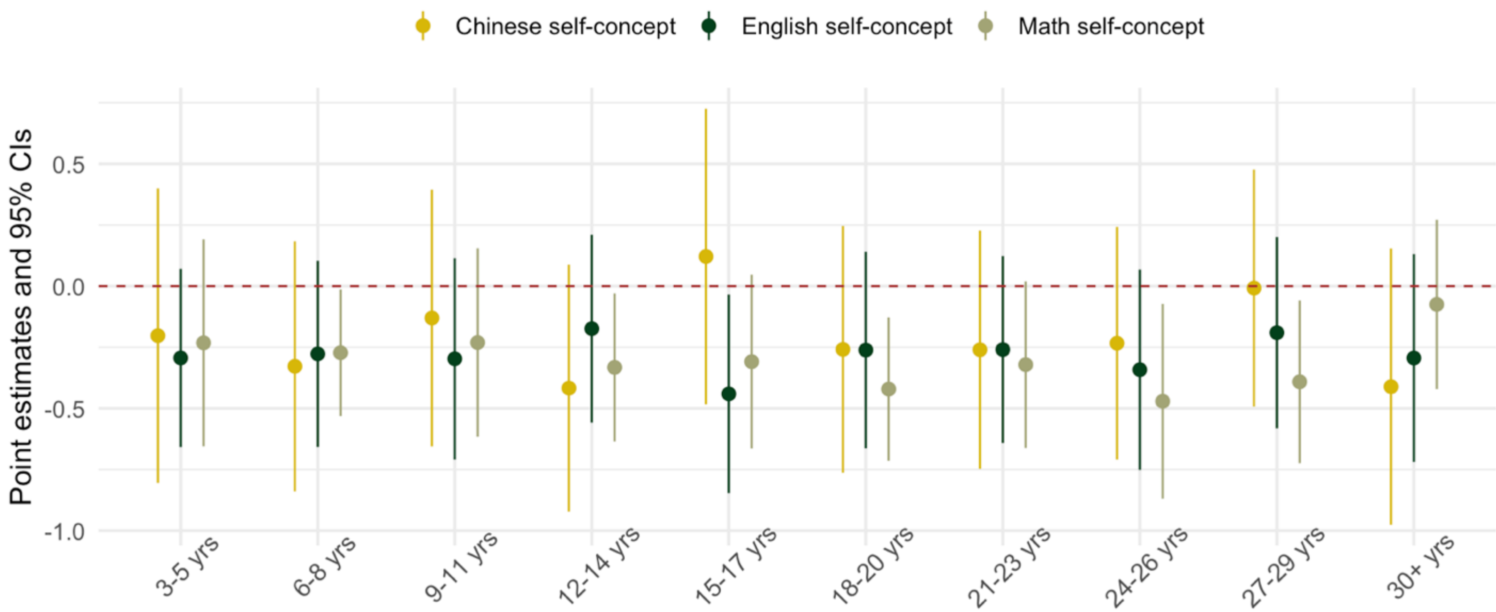



Figure 6. Experience on self-concept for schools without high turnover

*Notes*: Figure 6 demonstrates the average relationship between the listed bins of teacher experience and students' self-concept outcomes in Chinese, English, and math from an ordinary least squares regression estimator, which regresses each outcome on a set of indicators that each represents a year bin (0-2 years as reference level), after accounting for student-, homeroom-, and teacher-level covariates as well as school fixed-effects. Standard errors are clustered at school level. Note that the schools reporting high levels of teacher turnover are excluded.

Table C1. Comparing schools included and excluded from the main analyses on observed school characteristics

| School Characteristics | Included N = 63 | Excluded N = 49 | $p$-value |
|---|---|---|---|
| School district sampling frame | | | 0.069 |
| Sample frame 1 | 46.03% | 63.27% | |
| Sample frame 2 | 15.87% | 4.08% | |
| Sample frame 3 | 38.10% | 32.65% | |
| School district location | | | 0.03 |
| East China | 68.25% | 51.02% | |
| Middle China | 9.52% | 28.57% | |
| West China | 22.22% | 20.41% | |
| School district administrative level | | | 0.018 |
| Municipality | 28.57% | 12.24% | |
| Urban area of provincial capital cities | 20.63% | 14.29% | |
| Urban area of prefecture-level cities | 20.63% | 14.29% | |
| County or county-level city | 30.16% | 59.18% | |
| District population average education (years) | 9.88 (1.44) | 9.27 (1.34) | 0.024 |
| School location | | | 0.9 |
| Center of the city/town | 41.27% | 32.65% | |
| Outskirts of the city/town | 11.11% | 10.20% | |
| Rural-urban fringe zone of the city/town | 14.29% | 16.33% | |
| Towns outside of the city/town | 15.87% | 20.41% | |
| Rural areas | 17.46% | 20.41% | |
| Proportion of rural residency students | | | 0.003 |
| Lower than 25% | 33.33% | 8.16% | |
| 25% to 60% | 30.16% | 22.45% | |
| 60% to 80% | 15.87% | 30.61% | |
| Higher than 80% | 20.63% | 38.78% | |
| Proportion of the local students | | | 0.009 |
| Lower than 50% | 4.76% | 14.29% | |
| 50% to 70% | 26.98% | 12.24% | |
| 70% to 90% | 34.92% | 18.37% | |
| higher than 90% | 33.33% | 55.10% | |
| Number of substitute teachers | 1.38 (3.77) | 4.27 (17.86) | 0.5 |
| Unknown | 3 | 4 | |
| Average homeroom size | 48 (9) | 52 (8) | 0.011 |

*Notes:* Cells report mean and standard deviation for continuous variables and percentage of each category for categorical variables. The $p$-statistic was obtained from a) Wilcoxon rank sum test for district population average education, number of substitute teachers, and average homeroom size, and b) Pearson's Chi-squared test for all other characteristics.

Table C2. Heterogeneity in teacher education effects

Panel A. Students from low-income families

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Education (yrs) | 0.050 | 0.039 | 0.037 | 0.026 | 0.019 | 0.025 | -0.042 | -0.037 | -0.040 |
| | (0.046) | (0.050) | (0.066) | (0.025) | (0.024) | (0.019) | (0.028) | (0.034) | (0.029) |
| Low income | -0.286 | -0.269 | -0.200 | -0.110 | -0.089 | -0.065 | -1.000 | -1.034 | -1.119 |
| | (0.814) | (0.824) | (0.792) | (0.371) | (0.338) | (0.353) | (0.620) | (0.637) | (0.665) |
| Education x Low income | 0.019 | 0.017 | 0.013 | 0.006 | 0.004 | 0.003 | 0.066 | 0.068 | 0.074 |
| | (0.051) | (0.052) | (0.050) | (0.024) | (0.022) | (0.023) | (0.039) | (0.040) | (0.042) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.598 | 0.601 | 0.601 | 0.701 | 0.702 | 0.703 | 0.608 | 0.609 | 0.612 |

Panel B. Students with rural residency

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Education (yrs) | 0.070 | 0.058 | 0.055 | 0.019 | 0.010 | 0.015 | -0.022 | -0.015 | -0.016 |
| | (0.060) | (0.059) | (0.071) | (0.026) | (0.025) | (0.021) | (0.030) | (0.035) | (0.032) |
| Rural | 0.632 | 0.598 | 0.635 | -0.399 | -0.421 | -0.424 | 0.126 | 0.149 | 0.152 |
| | (0.736) | (0.652) | (0.649) | (0.265) | (0.263) | (0.254) | (0.443) | (0.443) | (0.468) |
| Education x Rural | -0.040 | -0.037 | -0.040 | 0.024 | 0.026 | 0.026 | -0.007 | -0.008 | -0.009 |
| | (0.046) | (0.041) | (0.041) | (0.017) | (0.017) | (0.016) | (0.028) | (0.028) | (0.029) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.599 | 0.601 | 0.602 | 0.701 | 0.702 | 0.703 | 0.608 | 0.609 | 0.612 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where subject score is regressed against teacher education, low-income (or rural residency), and their interaction term. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

Table C3. Heterogeneity in teacher experience effects

Panel A. Students from low-income families

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Experience (yrs) | -0.001 | 0.001 | 0.002 | -0.003 | -0.001 | -0.003 | -0.004 | -0.005 | -0.004 |
| | (0.004) | (0.004) | (0.004) | (0.003) | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) |
| Low-income | 0.105 | 0.080 | 0.068 | 0.073 | 0.062 | 0.072 | 0.035 | 0.048 | 0.040 |
| | (0.074) | (0.080) | (0.078) | (0.070) | (0.075) | (0.074) | (0.063) | (0.066) | (0.065) |
| Experience x Low-income | -0.006 | -0.005 | -0.004 | -0.005 | -0.005 | -0.005 | 0.001 | 0.000 | 0.001 |
| | (0.004) | (0.005) | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.598 | 0.601 | 0.601 | 0.702 | 0.702 | 0.704 | 0.608 | 0.609 | 0.612 |

Panel B. Students with rural residency

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Experience (yrs) | -0.001 | 0.002 | 0.003 | -0.003 | -0.002 | -0.004 | -0.004 | -0.005 | -0.005 |
| | (0.004) | (0.004) | (0.005) | (0.003) | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) |
| Rural | 0.035 | 0.046 | 0.033 | 0.002 | 0.002 | 0.001 | -0.020 | 0.005 | -0.001 |
| | (0.063) | (0.055) | (0.053) | (0.037) | (0.036) | (0.036) | (0.060) | (0.060) | (0.058) |
| Experience x Rural | -0.002 | -0.003 | -0.002 | -0.001 | -0.001 | -0.001 | 0.002 | 0.001 | 0.001 |
| | (0.004) | (0.003) | (0.003) | (0.002) | (0.002) | (0.002) | (0.003) | (0.003) | (0.003) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.598 | 0.601 | 0.601 | 0.701 | 0.702 | 0.703 | 0.608 | 0.610 | 0.612 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where subject score is regressed against teacher experience, low-income (or rural residency), and their interaction term. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

Table C4. Nonlinear effects of teacher experience on student academic outcomes

Panel A. Subject score outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| 3-5 yrs | 0.169 | 0.170 | 0.170 | 0.001 | -0.046 | -0.137 | 0.046 | 0.011 | -0.019 |
| | (0.185) | (0.145) | (0.173) | (0.093) | (0.104) | (0.105) | (0.192) | (0.235) | (0.170) |
| 6-8 yrs | 0.023 | 0.109 | 0.120 | -0.177* | -0.226* | -0.293** | 0.002 | -0.075 | -0.084 |
| | (0.222) | (0.190) | (0.183) | (0.078) | (0.099) | (0.096) | (0.166) | (0.216) | (0.153) |
| 9-11 yrs | -0.158 | -0.069 | -0.041 | -0.177* | -0.182 | -0.274** | 0.099 | 0.032 | 0.014 |
| | (0.187) | (0.163) | (0.163) | (0.071) | (0.108) | (0.101) | (0.156) | (0.188) | (0.137) |
| 12-14 yrs | -0.235 | -0.177 | -0.173 | -0.024 | -0.024 | -0.137 | -0.072 | -0.122 | -0.129 |
| | (0.197) | (0.175) | (0.194) | (0.084) | (0.097) | (0.093) | (0.168) | (0.199) | (0.142) |
| 15-17 yrs | 0.184 | 0.321* | 0.415* | -0.050 | -0.130 | -0.217 | -0.010 | -0.055 | -0.032 |
| | (0.177) | (0.145) | (0.162) | (0.098) | (0.128) | (0.130) | (0.166) | (0.196) | (0.145) |
| 18-20 yrs | -0.117 | -0.008 | 0.032 | -0.118 | -0.146 | -0.255* | -0.071 | -0.107 | -0.164 |
| | (0.177) | (0.150) | (0.164) | (0.088) | (0.109) | (0.108) | (0.163) | (0.195) | (0.141) |
| 21-23 yrs | -0.012 | 0.057 | 0.045 | -0.195* | -0.231* | -0.319** | -0.067 | -0.125 | -0.152 |
| | (0.190) | (0.152) | (0.164) | (0.083) | (0.101) | (0.095) | (0.175) | (0.212) | (0.145) |
| 24-26 yrs | -0.030 | 0.032 | 0.134 | -0.250* | -0.249* | -0.336** | 0.008 | 0.003 | -0.080 |
| | (0.191) | (0.165) | (0.183) | (0.107) | (0.118) | (0.119) | (0.198) | (0.187) | (0.150) |
| 27-29 yrs | -0.080 | 0.047 | 0.104 | -0.123 | -0.110 | -0.225 | -0.122 | -0.204 | -0.182 |
| | (0.174) | (0.155) | (0.152) | (0.120) | (0.129) | (0.122) | (0.163) | (0.207) | (0.148) |
| 30+ yrs | -0.088 | -0.145 | -0.105 | 0.002 | -0.036 | -0.177 | 0.010 | -0.084 | -0.004 |
| | (0.182) | (0.220) | (0.246) | (0.096) | (0.119) | (0.126) | (0.171) | (0.193) | (0.146) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.598 | 0.601 | 0.601 | 0.701 | 0.702 | 0.703 | 0.608 | 0.610 | 0.612 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where subject score is regressed against the categorical predictor: teacher experience measured in year bins (0-2 years omitted). Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

Panel B. Subject self-concept outcomes

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| 3-5 yrs | -0.292 | -0.194 | -0.299 | -0.005 | -0.115 | -0.317* | 0.314 | 0.215 | 0.190 |
| | (0.247) | (0.244) | (0.231) | (0.139) | (0.118) | (0.136) | (0.314) | (0.320) | (0.306) |
| 6-8 yrs | -0.541* | -0.403 | -0.427 | -0.043 | -0.152 | -0.300* | 0.189 | 0.039 | 0.051 |
| | (0.263) | (0.246) | (0.239) | (0.133) | (0.128) | (0.127) | (0.287) | (0.269) | (0.285) |
| 9-11 yrs | -0.358 | -0.164 | -0.195 | -0.010 | -0.103 | -0.304* | 0.315 | 0.212 | 0.217 |
| | (0.238) | (0.230) | (0.230) | (0.110) | (0.127) | (0.132) | (0.314) | (0.312) | (0.309) |
| 12-14 yrs | -0.507* | -0.358 | -0.458 | 0.086 | 0.046 | -0.202 | 0.172 | 0.077 | 0.090 |
| | (0.230) | (0.222) | (0.236) | (0.121) | (0.123) | (0.134) | (0.292) | (0.273) | (0.279) |
| 15-17 yrs | -0.301 | -0.203 | -0.090 | -0.025 | -0.272 | -0.465** | 0.101 | 0.028 | 0.059 |
| | (0.230) | (0.227) | (0.224) | (0.133) | (0.148) | (0.146) | (0.300) | (0.291) | (0.295) |
| 18-20 yrs | -0.504* | -0.331 | -0.359 | 0.086 | -0.053 | -0.295* | 0.086 | -0.043 | -0.062 |
| | (0.231) | (0.226) | (0.234) | (0.133) | (0.133) | (0.135) | (0.289) | (0.272) | (0.274) |
| 21-23 yrs | -0.412 | -0.266 | -0.381 | 0.024 | -0.096 | -0.290* | 0.161 | 0.112 | 0.118 |
| | (0.231) | (0.228) | (0.231) | (0.124) | (0.120) | (0.127) | (0.298) | (0.290) | (0.292) |
| 24-26 yrs | -0.566* | -0.427 | -0.307 | -0.150 | -0.184 | -0.376* | 0.033 | 0.072 | 0.039 |
| | (0.274) | (0.280) | (0.267) | (0.148) | (0.149) | (0.144) | (0.310) | (0.321) | (0.307) |
| 27-29 yrs | -0.392 | -0.159 | -0.153 | 0.086 | 0.014 | -0.238 | 0.043 | -0.037 | -0.011 |
| | (0.211) | (0.234) | (0.225) | (0.166) | (0.140) | (0.142) | (0.288) | (0.285) | (0.290) |
| 30+ yrs | -0.465* | -0.319 | -0.475 | 0.160 | -0.029 | -0.339* | 0.279 | 0.266 | 0.349 |
| | (0.181) | (0.231) | (0.267) | (0.139) | (0.148) | (0.154) | (0.301) | (0.286) | (0.302) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.079 | 0.083 | 0.087 | 0.255 | 0.258 | 0.261 | 0.241 | 0.245 | 0.246 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where subject self-concept is regressed against the categorical predictor: teacher experience measured in year bins (0-2 years omitted). Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender and indicator of advisor role. All models control for school fixed effects and cluster standard errors at school level.

## REFERENCES CITED


Aaronson, Barrow, L., & Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. Journal of Labor Economics, 25(1), 95–135. https://doi.org/10.1086/508733

Blazar, & Kraft, M. A. (2017). Teacher and Teaching Effects on Students' Attitudes and Behaviors. Educational Evaluation and Policy Analysis, 39(1), 146–170. https://doi.org/10.3102/0162373716670260

Boyd, Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. Journal of Policy Analysis and Management, 27(4), 793–818. https://doi.org/10.1002/pam.20377

Chen. (2020). A historical review of professional learning communities in China (1949-2019): some implications for collaborative teacher professional development. Asia Pacific Journal of Education, 40(3), 373–385. https://doi.org/10.1080/02188791.2020.1717439

Chetty, Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. The American Economic Review, 104(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Chu, Loyalka, P., Chu, J., Qu, Q., Shi, Y., & Li, G. (2015). The impact of teacher credentials on student achievement in China. China Economic Review, 36, 14–24. https://doi.org/10.1016/j.chieco.2015.08.006

Clotfelter, Ladd, H. F., & Vigdor, J. L. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. The Journal of Human Resources, XLI(4), 778–820. https://doi.org/10.3368/jhr.XLI.4.778

Clotfelter, Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review, 26(6), 673–682. https://doi.org/10.1016/j.econedurev.2007.10.002

Coenen, Cornelisz, I., Groot, W., Maassen van den Brink, H., & Van Klaveren, C. (2018). Teacher Characteristics and Their Effects on Student Test Scores: A Systematic Review: Teacher Characteristics and Their Effects on Student Test Scores. Journal of Economic Surveys, 32(3), 848–877. https://doi.org/10.1111/joes.12210

Coleman. (1966). Equality of educational opportunity. U.S. Govt. Print. Off.

Eble, & Hu, F. (2020). Child beliefs, societal beliefs, and teacher-student identity match. Economics of Education Review, 77, 101994. https://doi.org/10.1016/j.econedurev.2020.101994

Gong, Lu, Y., & Song, H. (2018). The effect of teacher gender on students' academic and noncognitive outcomes. Journal of Labor Economics, 36(3), 743–778. https://doi.org/10.1086/696203

Guarino, Dieterle, S. G., Bargagliotti, A. E., & Mason, W. M. (2013). What Can We Learn About Effective Early Mathematics Teaching? A Framework for Estimating Causal Effects Using Longitudinal Survey Data. Journal of Research on Educational Effectiveness, 6(2), 164–198. https://doi.org/10.1080/19345747.2012.706695

Hanushek. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. Journal of Economic Literature, 24(3), 1141–1177.

Hanushek. (2011). The economic value of higher teacher quality. Economics of Education Review, 30(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

Hanushek, & Rivkin, S. G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. The American Economic Review, 100(2), 267–271. https://doi.org/10.1257/aer.100.2.267

Harris, & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. Journal of Public Economics, 95(7), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Hu, Wang, S., Song, Y., & Roberts, S. K. (2022). Profiles of Provision for Learning in Preschool Classrooms in Rural China: Associated Quality of Teacher-child Interactions and Teacher Characteristics. Early Education and Development, 33(1), 121–138. https://doi.org/10.1080/10409289.2020.1802567

Kane, Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness?: Evidence from New York City. Economics of Education Review, 27(6), 615–631. https://doi.org/10.1016/j.econedurev.2007.05.005

Kim, Cho, H., & Kim, L. Y. (2019). Socioeconomic Status and Academic Outcomes in Developing Countries: A Meta-Analysis. Review of Educational Research, 89(6), 875–916. https://doi.org/10.3102/0034654319877155

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. Journal of Human Resources, 54(1), 1-36.

Ladd, & Sorensen, L. C. (2017). Returns to Teacher Experience: Student Achievement and Motivation in Middle School. Education Finance and Policy, 12(2), 241–279. https://doi.org/10.1162/EDFP_a_00194

Liu, & Hallinger, P. (2018). Principal Instructional Leadership, Teacher Self-Efficacy, and Teacher Professional Learning in China: Testing a Mediated-Effects Model. Educational Administration Quarterly, 54(4), 501–528. https://doi.org/10.1177/0013161X18769048

Liu, Peng, P., & Luo, L. (2020). The Relation Between Family Socioeconomic Status and Academic Achievement in China: A Meta-analysis. Educational Psychology Review, 32(1), 49–76. https://doi.org/10.1007/s10648-019-09494-0

Lockwood, & McCaffrey, D. F. (2014). Correcting for Test Score Measurement Error in ANCOVA Models for Estimating Treatment Effects. Journal of Educational and Behavioral Statistics, 39(1), 22–52. https://doi.org/10.3102/1076998613509405

Marsh, Kong, C.-K., & Hau, K.-T. (2001). Extension of the Internal/External Frame of Reference Model of Self-Concept Formation: Importance of Native and Nonnative Languages for Chinese Students. Journal of Educational Psychology, 93(3), 543–553. https://doi.org/10.1037/0022-0663.93.3.543

Möller, Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. Review of Educational Research, 79(3), 1129–1167. https://doi.org/10.3102/0034654309337522

Papay, & Kraft, M. A. (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, 105–119. https://doi.org/10.1016/j.jpubeco.2015.02.008

Rice. (2013). Learning from Experience? Evidence on the Impact and Distribution of Teacher Experience and the Implications for Teacher Policy. Education Finance and Policy, 8(3), 332–348. https://doi.org/10.1162/EDFP_a_00099

Rockoff. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. The American Economic Review, 94(2), 247–252. https://doi.org/10.1257/0002828041302244

Sun, Shafiq, M. N., McClure, M., & Guo, S. (2020). Are there educational and psychological benefits from private supplementary tutoring in Mainland China? Evidence from the China Education Panel Survey, 2013–15. International Journal of Educational Development, 72, 102144. https://doi.org/10.1016/j.ijedudev.2019.102144

Wayne, & Youngs, P. (2003). Teacher characteristics and student achievement gains: a review. Review of Educational Research, 73(1), 89–122. https://doi.org/10.3102/00346543073001089

Xu, D., & Li, Q. (2018). Gender achievement gaps among Chinese middle school students and the role of teachers' gender. Economics of Education Review, 67, 82-93.

Xu, Zhang, Q., & Zhou, X. (2022). The Impact of Low-Ability Peers on Cognitive and Noncognitive Outcomes: Random Assignment Evidence on the Effects and Operating Channels. The Journal of Human Resources, 57(2), 555–596. https://doi.org/10.3368/jhr.57.2.0718-9637R2

Zhao, Ye, J., Li, Z., & Xue, S. (2017). How and why do Chinese urban students outperform their rural counterparts? China Economic Review, 45, 103–123. https://doi.org/10.1016/j.chieco.2017.06.006

# CHAPTER III. CAUSAL IMPACTS OF TEACHER-STUDENT RELATIONSHIPS ON STUDENT ACADEMIC OUTCOMES, AND THE ROLE OF TEACHER-ADVISORS

## Introduction

The benefits of teacher-student relationships (TSRs) are well theorized from attachment, motivation, and sociocultural perspectives (Davis, 2003; Roorda et al., 2011). Researchers have long observed that positive TSRs are associated with reduced disciplinary problems (Crosnoe et al., 2004; Quin, 2017), enhanced psychological engagement and school attachment (Quin, 2017), and increased academic performance (Lee, 2012; Roorda et al., 2011). The overall correlations between positive TSRs and better student outcomes were found to be medium to large in several systematic literature reviews (e.g., Christensen, 1960; Cornelius-White, 2007; Quin, 2017; Roorda et al., 2011). Not surprisingly, many researchers, policymakers, educators, and parents hold a fairly strong intuition that "Kids don't learn from people they don't like" (Aspy et al., 1977) and believe affective education increases school productivity (Aspy and Roebuck, 1982).

From a policy and economic perspective, if teacher-student relationships do have positive, causal effects on student performance, it will be one of the most cost-effective ways to boost school outcomes because most of the commonly implemented relationship-building policy initiatives such as integrating advising or mentoring responsibilities into teachers' roles, teacher preparation or on-site professional development trainings focusing on interpersonal skills and social-emotional learning strategies, introducing relationship measures into teacher evaluation, etc., usually require fewer resources than traditional school reform initiatives such as hiring high-performing building leaders or teachers, and reducing class sizes, student-teacher ratios, or student-counselor ratios.

Unfortunately, existing literature on TSR is overwhelmingly correlational. For example, Roorda et al. (2011) reviewed 99 studies and found that most of them used cross-sectional designs and none supported causal inference. The same pattern was documented in the other two literature review studies, Cornelius-White (2007) and Quin (2017). Compared to correlational studies, experimental evidence on the effects of TSR is extremely scarce (Moore et al., 2019). To my knowledge, only a few small-scale, experimental studies provided some potential channels through which TSR may influence student learning. For example, mentoring supports from teachers (e.g., setting academic goals, developing learning strategies, progress reviews, and positive feedback) were found to significantly improve student academic outcomes such as GPA at school (Murray & Malmgren, 2005).

Indeed, whether and to what extent TSRs affect student academic performance is a challenging research question: it cannot be answered by directly comparing the average performance of students who have positive relationships with their teachers and that of who do not, because unobserved differences between these two groups introduce bias into the estimation. For instance, highly motivated or socioeconomically advantaged students often form more positive relationship with their teachers; if they can successfully sort to teachers before or during the school year, the differences in academic outcomes between these students and their peers are attributable to their motivation or family wealth (and potentially more confounding factors) rather than TSR. In addition to confounding issues, reverse causality also creates substantial threats to internal validity; if high performing students tend to form more positive relationship with teachers, the observed relationships between TSR and student outcomes will be contaminated by the effects from student performance to TSR.

Education experiments overcome these methodological barriers by randomly assigning teachers and students to different levels of TSR but face incredible practical and ethical issues; thus, to my knowledge, have never been done at scale in any countries. I venture into this frontier by implementing a quasi-experimental design, where different levels of TSR are defined by an exogenous shock rather than the self-selection of teachers and students. This approach is arguably one of the best alternatives to education experiments and importantly, is possible to conduct at-scale with observational data.

Specifically, I leverage a natural experimental condition in Chinese junior high schools (grades 7-9) where students were randomly assigned to teachers upon their entry to 7th grade and some of them were randomly assigned to a dual-role teacher-advisor (a teacher that not only teaches a core content subject but also formally serves as the student's advisor). Based on extended attachment theory (Birch & Ladd, 1997; Pianta, 1999) and social-emotional learning literature (Elias et al., 1997), being taught by advisor may positively affect student's relationship with teacher, and, since it is a condition randomly assigned to student, it may serve as an ideal instrumental variable (IV) to identify an exogenous portion of variance in TSR then further identify the causal relationship from TSR to student outcomes. In other words, using an instrumental variable estimation (IVE; e.g., Angrist et al., 1996; Angrist & Krueger, 2001; Baiocchi et al., 2014; Bound et al., 1995; Staiger & Stock, 1997) approach, I am able to identify TSR effects without explicitly controlling for numerous omitted variables that also influence student outcomes.

In this large-scale, longitudinal, quasi-experimental study, I answer two main research questions: (1) whether—and to what extent—being taught by a teacher-advisor affects teacher-student relationships and academic outcomes in China? and (2) whether—and to what extent—teacher-student relationships affect student academic outcomes in China?

## Literature Review

### Teacher-Student Relationship and Student Learning

The theoretical framework for the TSR effects on students learning is well-developed in the literature. The extended attachment theory literature (Birch & Ladd, 1997; Pianta, 1999) posits teachers as potential attachment figures to students at school (Rhodes et al., 2006). The central idea of attachment theory is that children's emotional safety toward their mother allows them to explore their environment and develop social and cognitive competencies (Bowlby, 1969). Extending this mechanism into a school setting, it is expected that if a similar emotional bond is established between teachers and students, students will build confidence and motivation, become more engaged in learning activities, and actively develop academic skills (Birch & Ladd, 1997; Roorda et al., 2011; Pianta et al, 2012). Furthermore, the social-emotional learning literature also sheds light on one of the mechanisms through which TSR may impact student learning. A positive learning environment will help students become social-emotionally competent (Elias et al., 1997) and students' emotional attachment to school and engagement in classroom are critical components that influence student performance (Becker & Luthar, 2002; Hoffman, 2009).

The close relationship between TSR and student outcomes has been long-observed and well-documented by researchers. There have been two literature review studies since Davis (2003) highlighted the strong associations between TSR and student social and cognitive development. Cornelius-White's (2007) meta-analysis on 119 papers from six nations identified an overall substantial, positive association between teacher-student relationship and student cognitive outcomes (e.g., achievement, perceived achievement, grades, IQ, etc.). Roorda et al (2011) conducted a meta-analysis of 92 articles from five regions and nations and found that the positive relationship between teachers and students was positively associated with students' school engagement and achievement. Although few of these three literature reviews and the studies they synthesized were able to establish a causal linkage from TSR to student outcome, they contributed to our current understanding of TSR and inspired various policy initiatives at local, national, and international level to use TSR as a leverage to boost student outcomes.

**Teacher Advising and Student Learning**

Given the suggestive evidence on the potential benefits of positive teacher-student relationships and separate evidence on the potential of guidance counseling for secondary students (Carrell & Hoekstra, 2014; Hurwitz and Howell 2014; Mulhern, 2020), many school systems across the world have implemented some form of an advisory program. While advisory programs assume many forms, they generally follow a format in which an advisor is assigned to a small group of students to provide individualized support on students' academic and personal developments (Galassi et al., 1997; McClure et al., 2010). In classroom-based school systems (e.g., the U.S.) where students are selected to teachers based on students' own schedule, a student's advisor may or may not be their classroom teacher. In homeroom-based school systems (e.g., China) where students are grouped in homerooms and share a common homeroom schedule, a homeroom's advisor is one of their core content teachers. Whereas advisory programs vary considerably across school settings and even schools under the same setting, they share the common theory of change – teacher advising is correlated with improved TSR and enhanced student outcomes.

In the U.S., advisory programs have been an evidence-based junior high school movement starting from the past century. Some states and education agencies leveraged various policies to reenforce the teacher advisement implementation, for example, Florida passed legislation in the 1980s to fund junior high school advisory programs (Galassi et al., 1997). Advisory programs had also been advocated by the National Association of Secondary School Principals (NASSP) and been integrated into multiple school reform efforts such as the Model Schools Project, IDEA's Individually Guided Education, and the Reform in Secondary Education Project in California. To date, teacher advising still serves as a complement to the school counseling system in many schools across the nation.

Unlike the school counseling programs implemented by systematically trained professionals under national or state guidelines, however, school advisory programs vary considerably from school to school, even under the same education system and settings. Subject to local policy, advisory programs may be designed to meet one or multiple student needs such as personal advocacy, group identity, development guidance, invigoration, academic performance, and general school business (Galassi et al., 1997), and teacher-advisors' responsibilities are difficult to universally define or clearly categorize. As a result, advisory program evaluation studies suffer from the substantial lack of uniformity across advisory programs, data limitation due to the small scopes of implementation, and methodological weaknesses in addressing omitted variable bias created by the prevalent sorting between advisors and students (Galassi et al., 1997). It is no surprise that advisory is an extremely understudied area and that within the scarce literature, findings are considerably inconclusive and mixed (Galassi et al., 1997; McClure et al., 2010) with only very few, small-scale experimental studies highlighting that supports from teachers (e.g., setting academic goals, developing learning strategies, progress reviews, and positive feedback) significantly improve student academic outcomes such as GPA at school (Murray & Malmgren, 2005).

The existing literature on teacher advising conducted in Chinese education context is even more limited. In Wang and Yang (2021), the most relevant work to this paper, the authors exploited the random teacher-student assignment in the data and found that "homeroom teachers" (or teacher-advisors in the current study) had positive impact on student academic outcomes (measured by a pooled raw score in three core subjects, Chinese, English, and math). In their exploratory analysis, they found that classroom teacher-student relationships might be the potential mechanism through which these advisor effects generated. In other words, Wang and Yang (2021) add evidence to the theory of change underlying advisory programs: teacher advising is reliably correlated with improved teacher-student relationship and enhanced student outcomes, which provide supportive information to the assumptions I make in my IV estimation. Another relevant but more peripheral study, Chen and Zhao (2022), found that the administrative duties of homeroom teachers, such as grade-level instruction team leadership and department head significantly curbed student performance.

## Background and School Settings

I conduct my research in China and identify my population of interest as Chinese public junior high school (grades 7-9) students and their core content teachers based on a critical policy consideration: Chinese public junior high schools are under a national law that drives the implementation of random assignment of teachers to students. To contextualize this natural experiment, it is helpful to note that China and many other countries such as France, Germany, India, Israel, Japan, and South Korea share a homeroom-based school system, which is different than the classroom-based settings in the U.S., UK, and many Western countries. Specifically, unlike in the U.S. where each student has their own schedule and attends different classrooms each school day, Chinese students are grouped into homerooms, put on a shared homeroom schedule, and assigned a group of subject teachers who rotate to the homeroom to teach.

Among the group of subject teachers, one of them formally serves as the homeroom's advisor (or teacher-advisor, see Galassi et al, 1997 for more interchangeably used terms for this role). As a result, a natural comparison condition is formed between students who are taught by traditional teachers and those taught by teacher-advisors, in a given content subject, e.g., math. To fully understand the widespread presence of teacher-advisors, it is helpful to know that school guidance counselor is not a professional position in China, therefore in compensating for this policy void, every homeroom in the country has a formal teacher-advisor in position to implement a comprehensive advisory program. According to the Ministry of Education's regulations of teacher-advisors in 2009 (henceforth referred to as 2009 Regulation; See Appendix B2 from the previous chapter for more details), these advisory programs generally integrate four core components: moral education, student discipline, student development, and mentoring. Acting on these responsibilities and utilizing the weekly advisory periods as opportunity of social-emotional learning, teacher-advisors play a significant role in students' school life and often form much closer relationships with advisees than traditional teachers do.

Throughout all years in which students attend the same school, students typically remain grouped with their original homeroom cohorts. Their teacher-advisor and core content teachers (especially in subject areas that require three-year curriculums) are encouraged to follow the homerooms rising to higher grades to gain familiarity of the full junior high curricula and teaching materials. This is particularly true for teachers who teach Chinese (language arts), English (nationally mandated foreign language), and math – the only three core subjects that not only require a full three-year education but also have larger weights over other subjects (such as history, political studies, physics, chemistry, geometry, biology, etc.) in the high-stakes high-school entry exam upon students' graduation. In conjunction, although folk knowledge suggests that all subject teachers are expected to serve on this role when they are needed and their personal situation allows them to, the majority of teacher-advisors are Chinese, English, and math teachers so that it will be convenient for them to follow the homeroom cohort rising to higher grades and maintain a stable homeroom ecosystem.

**Natural Experiment Background**

In 2006, with great attention to education equality, the Compulsory Education Law (henceforth referred to as the 2006 Law; see Appendix B1 for more details) called off student tracking at all compulsory education levels (grades 1-9) and effectively eliminated national-, province-, and district-level academic exams below grade 9. The 2006 Law has stimulated a trend of random assignment of teachers to students across the nation, which was captured seven years later in the first nationally representative educational survey, the China Education Panel Survey (CEPS): 83 percent of the randomly sampled schools across the nation reported that they randomly assigned teachers to students upon students' entry to junior high school.

Both from the literature (e.g., Xu et al., 2022) and folk knowledge, the common teacher-student assignment approach has been that, supervised by local education departments, schools create either random or stratified homerooms of students upon students' entry to school, and then randomly assign teacher groups to homerooms (teachers are often assigned to multiple groups depending on their workload, for example, a math teacher is typically assigned to two homerooms because two classes per day, five days per week is the total full time equivalent workload for a junior high school math teacher). Adding to the validity of the random assignment, local education departments typically review their public schools every year to check whether there are violations of the 2006 Law. Their strategies vary but many may require schools to submit a copy of their original homeroom rosters for the purpose of documentation. Others may conduct a student and/or parent survey or conduct more detailed school reviews in occasions when parents complain about unlawful student tracking or kids being discriminated against during homeroom assignment. These policy regulations greatly reinforce the validity of random assignment and in turn help it become an educational norm accepted by students, parents, and educators across the nation. This random assignment is crucial to my identification strategy and more evidence will be presented in further details in the Method section.

## Data and Measures

The data information and sample restriction process for this article are largely identical with those in the previous chapter.

### Key Variables

**Predictor Variable.** To measure TSR, I retrieve three items from CEPS student survey that ask the student if their Chinese, English, or math teacher often praises them, asks questions of them, and pays attention to them in classroom. All items were rated on a 4-point Likert-scale ranging from "strongly disagree" to "strongly agree". I performed a principal component analysis (PCA) to uncover the construct(s) underlying the three items and present results in Appendix Table A1. Across three subjects, the first component was the only component with corresponding eigenvalue above one and explains more than 70 percent of the total variation in all three items. Given the robustness of the PCA results and the approximately equal weighting of all items, using first component or the average of the three items should not return different estimates. To promote interpretation, in my main analyses, I follow Garrett and Steinberg (2015) and take the mean of the three items to create a single index of TSR, then standardized it to be mean zero and unit variance within each subject, each school. I used the PCA first component measure of TSR in my robustness checks to confirm the findings in my main analysis.

**Instrumental Variable.** I use a single IV variable, a dichotomous variable coded one for students whose teacher was also their homeroom advisor and zero otherwise.

**Outcome Variables.** In each of the three subjects (Chinese, English, and math), student academic outcomes are measured by two variables, both of which contain unique information on student learning. First, I use students' subject-matter test score on school-administered mid-fall semester exam (obtained from students' school records). The second outcome variable is subject self-concept. The proxy I use is students' response to a 4-point Likert-scale survey item asking whether the subject is difficult. I reverse code the variable to represent four levels of self-concept: zero (very low), one (low), two (high), and three (very high). Note that both variables are standardized to be mean zero and unit variance within each school.

**Covariates.** I draw from wave 1 data three groups of covariates at the student-, homeroom-, and teacher-level to improve estimation precision. Student-level covariates include student wave 1 Chinese, English, math, and CEPS cognitive test scores as well as demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth (three categories including low-income, middle-class, and wealthy). The homeroom-level covariates include homeroom size and the homeroom means (leave-one-out mean, i.e., excluding self for each observation) of student characteristics. The teacher-level covariates include teacher gender, age, education attainment, and experience.

**Missing Data**

Within each of the analytic sample schools (N = 63), I matched students with core content teachers and obtained separate samples for Chinese, English, and math, then examined the missingness. The proportion of missingness on IV was 0.92% for Chinese sample and zero for the other two subjects. The predictor variable was missing at 0.79%, 1.06%, and 0.75% for Chinese, English, and math sample, respectively. Across three subjects, on outcome variables, standard score and self-concept, the range of missing rate is 1.08%-1.19% and 0.48%-0.54%. On student and teacher covariates, the missing rate is all below 2% except for three variables: student age, teacher age, and teaching experience, the highest missing rate is 2.42%, 2.25%, and 4.78%, respectively, across three subjects. Because of the relatively large sample size and small missingness, I assume this missingness is completely at random and drop all observations that have any missing value on IV, predictor, and outcome variables. I replace missing values on other variables with homeroom mean (for student covariates) or school mean (for teacher covariates).

The final sample size for Chinese, English, and math is 5,055, 5,080, and 5,105 students, respectively. Summary statistics of key variables are presented in Table 2.1. Note that the sample size and descriptive statistics of key variables are similar across three subjects with one exception, students seemed to be slightly more likely taught by teacher-advisor in math than in the other two subjects.

Table 2.1. Analytic sample summary statistics

| Key Variables | Chinese Sample | English Sample | Math Sample |
|---|---|---|---|
| | N = 5,055 | N = 5,080 | N = 5,105 |
| ***Predictor Variable*** | | | |
| TSR | 0.00 (1.00) | 0.01 (0.99) | 0.00 (0.99) |
| ***Instrumental Variable*** | | | |
| Taught by teacher-advisor | 28.11% | 27.64%% | 31.38% |
| ***Outcome Variables*** | | | |
| Score | 0.00 (0.99) | 0.00 (0.99) | 0.01 (0.99) |
| Self-concept | 0.00 (0.99) | 0.00 (0.99) | 0.00 (0.99) |

*Notes*: Cells report mean and standard deviation for continuous variables and percentage of each category for categorical variables.

## Method

### Identification Strategy

The exogenous variation in TSR is the key to my identification strategy and it relies on a critical assumption is that the assignment of teachers and students was random so that students taught by advisors and non-advisor teachers were not systematically different in wave 1. Indeed, the random assignment of students and teachers is not only enforced by the 2006 Law and reported by the surveyed schools (discussed in Background section), but also confirmed in the data. Specifically, I conduct a series of student covariates balance checks within each subject by regressing the instrumental variable, the indicator of being taught by teacher-advisor, against wave 1 student covariates while controlling for school fixed effects and clustering standard errors at school level. I show results in Table 2.2 and highlight that, both individually and jointly, students' wave 1 scores and other socioeconomic variables did not predict whether or not the student was assigned for their subject-matter class to a teacher-advisor.

The only exception is that students' same-subject wave 1 score was significantly correlated with being taught by an advisor, i.e., students having higher wave 1 math scores were significantly more likely taught by advisors who taught math, same as to Chinese and English.

Recall that wave 1 scores measure student performance on mid-fall semester exams, by that time, students had been in advisors' classroom for half a semester (2-3 months). Thus, one might conclude that this indicates the possibility of students with either better learning ability or stronger motivation of learning a subject, say, math, sorting to math teacher-advisors during that first half semester. I cannot rule out this possibility. However, noting the small magnitude of the coefficients and the insignificant coefficients of other variables that are also correlated with academic ability and motivation, I argue that these estimates are more likely capturing the potential effects of being taught by advisor for a short period of time (although unfortunately, without prior ability controls and careful modeling, these coefficients cannot be directly interpreted as teacher-advisor effects). Based on this reasoning, I argue that the critical assumption to my identification strategy was largely verified: students assigned to learn in subject classes with teacher-advisors were not systematically different from their school peers in observable ways.

Table 2.2. Covariates balance check: regressions of being taught by teacher-advisor on student wave 1 covariates

| | Being taught by teacher-advisor | | |
|---|---|---|---|
| | Chinese Sample | English Sample | Math Sample |
| Wave 1 Chinese | 0.026* | -0.013 | -0.013 |
| | (0.011) | (0.012) | (0.013) |
| Wave 1 English | -0.017 | 0.045** | -0.026 |
| | (0.015) | (0.015) | (0.017) |
| Wave 1 Math | -0.024 | -0.023 | 0.039* |
| | (0.013) | (0.014) | (0.016) |
| Wave 1 cognitive | -0.005 | 0.007 | 0.018 |
| | (0.008) | (0.016) | (0.016) |
| Female | -0.006 | -0.005 | 0.010 |
| | (0.009) | (0.007) | (0.009) |
| Age | 0.008 | -0.015 | 0.022 |
| | (0.011) | (0.012) | (0.016) |
| Rural residency | -0.011 | -0.004 | 0.000 |
| | (0.013) | (0.011) | (0.014) |
| Only child | -0.010 | 0.000 | -0.009 |
| | (0.009) | (0.012) | (0.012) |
| Migrant family | -0.006 | -0.009 | 0.014 |
| | (0.012) | (0.011) | (0.012) |
| Mother education (years) | 0.004 | -0.003 | 0.001 |
| | (0.002) | (0.002) | (0.002) |
| Father education (years) | 0.001 | 0.001 | 0.001 |
| | (0.002) | (0.002) | (0.002) |
| Family income | -0.007 | 0.010 | 0.013 |
| | (0.009) | (0.010) | (0.011) |
| School FE | X | X | X |
| School clustered SE | X | X | X |
| *F*-Statistics | 1.484<br>(df = 12; 62) | 1.416<br>(df = 12; 62) | 1.527<br>(df = 12; 62) |
| Observations | 5055 | 5080 | 5105 |
| $R^2$ | 0.541 | 0.598 | 0.466 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate ordinary least squares regression where the instrumental variable, the indicator of being taught by homeroom advisor, is regressed on wave 1 student score measures and characteristics. All models control for school fixed effects and cluster standard errors at school level.

**Instrumental Variable Estimation**

To overcome the endogenous issue related to my predictor variable TSR, I leverage the random assignment of teacher-advisors to students to identify a portion of variance in TSR that was uncorrelated with potential outcomes and use only this portion (rather than the endogenous TSR) to obtain asymptotically unbiased estimates of TSR effects on student academic outcomes. Intuitively, the IV, being taught by advisor, serves as “a haphazard push to accept a treatment where the push can affect the outcomes only to the extent that it alters the treatment received” (Yang et al., 2014) – therefore to statistically parse out the exogenous variation in TSR. Thus, my estimate of TSR effect is considered asymptotically unbiased under three critical assumptions: exogeneity, relevance, and exclusion restriction. The first assumption, exogeneity assumption – being taught by advisor was uncorrelated with the residuals in neither the regression equation of the reduced form (regression of outcome on IV) nor the first stage (regression of TSR on IV) – was warranted by my identification strategy discussed earlier.

To meet the relevance assumption, being taught by advisor must be (relatively) strongly correlated with TSR. Since advisor programs are designed to provide students individualized support and promote connections between advisors and advisees, students are expected to perceive more positive TSR in their advisors’ than in traditional teachers’ classroom. This expectation was confirmed in my data. Specifically, I show in Table 2.3 evidence that being taught by advisor significantly increased TSR by more than 0.2 standard deviations (SD)[1] in language subjects (Chinese and English). In other words, being taught by advisor meets the relevance assumption for an instrument in Chinese and English.

In the math sample, unfortunately, as shown in the last three columns of Table 2.3, the policy initiative of assigning advisors to students in hopes of improving TSR did not see the intended effects. This suggests that math teachers may influence students' academic outcomes through channels that beyond relationship building, which is potentially related to how Chinese students are tested in math: the nation does not allow any calculator usage in high-stakes high-school and college entry exams. Consequently, students often score higher by accurately perform complex hand calculations, they are used to improve their math scores by repeated practicing, and it is not uncommon that math teachers improve students' performance by assigning large amount of homework and overpreparing students for exams. Regardless these reasons, being taught by advisor did not meet the relevance assumption therefore could not serve as a valid IV in math subject. The causal impact of TSR on students' math outcomes remains unknown throughout this study.

---

1. Although this effect size is considered educationally meaningful, it is not statistically large enough for being taught by advisor to be a strong IV. One may notice from the 2SLS estimates in Table 2.6 that the first-stage F statistics, or Kleibergen-Paap rk Wald F statistics (Kleibergen & Paap, 2006) were all close to or below 10, indicating that being taught by advisor is a weak IV. However, since I only have one IV and one predictor, i.e., the number of instruments is equal to the number of endogenous predictor of interest, the bias of 2SLS regression is "approximately zero" (Angrist & Krueger, 2001). To conclude, although being taught by advisor is a weak instrument that may inflate the standard errors of the point estimates, it is unlikely to bias my 2SLS estimation, therefore I will proceed with this IV in Chinese and English samples

Table 2.3. Effects of being taught by teacher-advisor on teacher-student relationship

| | Teacher-Student Relationship | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Taught by advisor | 0.202** | 0.218** | 0.211** | 0.130 | 0.182** | 0.208** | 0.037 | 0.037 | 0.025 |
| | (0.072) | (0.082) | (0.077) | (0.072) | (0.063) | (0.069) | (0.079) | (0.076) | (0.074) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 | 5105 | 5105 | 5105 |
| R2 | 0.019 | 0.024 | 0.025 | 0.038 | 0.042 | 0.045 | 0.009 | 0.013 | 0.016 |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where the predictor of interest, teacher-student relationship, is regressed against the instrumental variable, the indicator of being taught by teacher-advisor. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and cluster standard errors at school level.

The last assumption, exclusion restriction, requires that being taught by advisor only impacts student learning through TSR. This assumption is known to be "untestable" and may be challenged if having a teacher-advisor systematically affected student outcomes without altering TSR. The biggest threat to this assumption is selection to teacher-advisor role, in other words, if schools consistently assigned more effective teachers to be homeroom advisors, then the IV estimates of the TSR effects on student performance will be biased. I recognize that directly testing this assumption is nearly impossible due to the fact that not all the homerooms were selected within each school. Acknowledging this limitation and the fact that two homerooms were randomly selected from each school and all schools were randomly selected from the nation, I reasonably relax this assumption to be that teacher-advisors and non-advisor teachers are not systematically different on variables that are associated with student academic outcomes.

This assumption is likely warranted based on two main reasons. Firstly, the 2009 Regulation only requires advisors to have mentoring and communication skills (see Appendix B2 of last chapter) – which are not pedagogical nor content specific, and folk knowledge among educators suggests that advisor appointment decision is finalized by teachers themselves and often hinges on their availability (e.g., health and family conditions) and willingness to take on such a committed role. Moreover, a wave 1 teacher characteristics balance check (Appendix Table A2) shows that, in my analytic sample, there was no systematic difference between 104 advisors and 251 non-advisor teachers in terms of pre-existing characteristics such as gender, age, teaching experience, education attainment, and subject area. Although this analysis only demonstrates that these two teacher groups do not differ based on observables and cannot rule out the possibility that the two groups may differ in unobserved ways, it makes this difference less plausible.

There may be another potential threat to the exclusion restriction if teacher-advisors impact students' performance by improving parents' involvement and investment in their children's education without altering classroom teacher-student relationships. Research has documented the positive relationship between parental involvement and student academic achievements (Fan & Chen, 2001) and home-based parental involvements such as family resources and study aids are more common in an Asian society (Ho, 2003). Due to the advisor role, teacher-advisors often have more direct and frequent communications with parents than traditional teachers do and close advisor-parent relationships are not uncommon in China. The most frequently observed, academically oriented changes in students – such as students who exhibit misbehaviors or have attention issues behave differently in their advisor's classroom or seek after-school tutoring on the subject taught by their advisor – could be results from improved teacher-student relationships (which would not violate the exclusion restriction assumption) but would do so if these changes were due to parental influence.

I argue that these potential backdoors are likely not posing a big threat because these students with disruptive behaviors, attention issues, or access to after-school tutoring are systematically different from other students in terms of wave 1 academic performance, cognitive score, and demographical characteristics – all of which have been adjusted for in my estimation. Nonetheless, I conduct various formats of robustness check and show evidence that findings in my main analysis hold after accounting for the fixed effects of 1) whether or not the student reported high frequency of at least one of the three at-risk factors including unable to concentrate, skipping classes/being absent/ truanting, and copying homework from others/cheating in exams; and 2) whether or not the student had off-school tutoring on the estimated subject.

Based on my identification strategy and estimation approach, I recover the impacts of TSR on student outcomes by estimating a value-added, two-stage least-squares (2SLS) regression in the following:

$$TSR_{ijt} = \alpha_g\,(g\,(A_{i,t-1})) + \alpha_1 ADVISOR_{jt} + \alpha_2 X_{i,t-1} + \alpha_3 P_{j,t-1} + \alpha_4 T_{j,t-1} + \theta_s + g_i \quad (1)$$

$$A_{it} = \beta_g\,(g\,(A_{i,t-1})) + \beta_1 \widehat{TSR}_{ijt} + \beta_2 X_{i,t-1} + \beta_3 P_{j,t-1} + \beta_4 T_{j,t-1} + \theta_s + \varepsilon_i \quad (2)$$

where i, j, s, t denote student, teacher (homeroom), school, year. In equation (1), $TSR_{ijt}$ is the endogenous predictor variable measuring teacher-student relationship and $ADVISOR_{jt}$ is the IV, the indicator of teacher-advisor. In equation (2), $A_{it}$ is student i's academic performance or self-concept in Chinese or English in year t and $\widehat{TSR}_{ijt}$ is the predicted value of teacher-student relationship by equation (1). In both equations, $g\,(A_{i,t-1})$ is the cubic functions of student i' s prior score in Chinese, English, math, and CEPS cognitive test, $X_{i,t-1}$, $P_{j,t-1}$, and $T_{j,t-1}$ are wave 1 student-, homeroom peer-level, and teacher-level covariates in year t-1, and $g_i$ and $\varepsilon_i$ are the idiosyncratic error terms. The coefficient of interest is $\beta_1$, which is the estimated effect of teacher-student relationship on student outcome. Note that I control for school fixed effects ($\theta_s$) to account for school time-invariant characteristics that include both students' and teachers' sorting to schools, then cluster standard errors at the school level to account for the within-school correlations among residuals. I estimate each of the two subjects, Chinese and English separately.

## Results

### Effects of Taught by Advisor on Student Outcomes

**Advisor Effects on TSR (First-Stage Estimation).** As discussed in the Method section, Table 2.3 demonstrates the first-stage of the 2SLS estimates: being taught by teacher-advisor for a school year (from wave 1 to wave 2) significantly improved TSR in Chinese and English by

0.211 standard deviation (SD) and 0.208 SD. This indicates that teacher-advisors have the intended positive effects on student social-emotional aspect of learning therefore being taught by advisor can be used as an instrument to identify the exogenously defined relationship between TSR and student outcomes in these two language subject areas (but not math).

**Advisor Effects on Academic Outcomes (Reduced-Form Estimation).** I present in Tables 2.4 the effects of being taught by a teacher-advisor on student academic outcomes. Specifically, being taught by advisor significantly improved student English and math score by 0.139 SD and 0.136 SD and increased student self-concept in Chinese and English by 0.216 SD and 0.178 SD. However, in Chinese test score and math self-concept, the advisor effects were not significantly different from zero. These findings were robust across three different model specifications. In Appendix Table A3, I used wave 2 CEPS cognitive test score as alternative outcome to re-estimate teacher-advisor effect and show evidence that my models did not have overidentification issues.

Note that compared to Wang and Yang (2021), where the authors used only wave 1 CEPS data and pooled information from three subjects to find positive effects of being assigned to teacher-advisor's classroom on both score and self-concept outcomes, my findings add to the literature in at least two ways. First, I add important heterogeneity estimates across different subjects and suggest that the mechanisms underlying advisor effects may be different for math and Chinese. Second, I estimate advisor effects in a value-added model and by accounting for student prior year achievements, I not only improve estimation precision but also allow for an expanded interpretation of my results – the effects of being taught by advisors for a full school year.

Table 2.4. Effects of being taught by advisor on student academic outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Taught by advisor | 0.050 | 0.042 | 0.043 | 0.100** | 0.133*** | 0.139*** | 0.114** | 0.119** | 0.136** |
| | (0.048) | (0.054) | (0.053) | (0.032) | (0.036) | (0.040) | (0.041) | (0.042) | (0.039) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 | 5105 | 5105 | 5105 |
| R2 | 0.598 | 0.600 | 0.601 | 0.704 | 0.705 | 0.705 | 0.608 | 0.610 | 0.611 |

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Taught by advisor | 0.222** | 0.218** | 0.216** | 0.161** | 0.208*** | 0.178*** | 0.056 | 0.038 | 0.016 |
| | (0.067) | (0.076) | (0.077) | (0.052) | (0.044) | (0.045) | (0.050) | (0.046) | (0.040) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 | 5105 | 5105 | 5105 |
| R2 | 0.081 | 0.085 | 0.086 | 0.261 | 0.265 | 0.265 | 0.237 | 0.240 | 0.245 |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where the outcome variable (score or self-concept), is regressed against the instrumental variable, the indicator of being taught by teacher-advisor. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and cluster standard errors at school level.

Table 2.5. Naïve estimates of the relationship between TSR and student academic outcomes

| | Subject Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| TSR | 0.053*** | 0.054*** | 0.054*** | 0.089*** | 0.089*** | 0.089*** | 0.051*** | 0.050*** | 0.049*** |
| | (0.011) | (0.011) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.599 | 0.602 | 0.602 | 0.701 | 0.702 | 0.703 | 0.608 | 0.609 | 0.612 |

| | Subject Self-Concept | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| TSR | 0.200*** | 0.202*** | 0.202*** | 0.195*** | 0.194*** | 0.194*** | 0.161*** | 0.161*** | 0.158*** |
| | (0.016) | (0.016) | (0.016) | (0.017) | (0.017) | (0.017) | (0.017) | (0.017) | (0.016) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 4882 | 4882 | 4882 | 5010 | 5010 | 5010 | 4995 | 4995 | 4995 |
| R2 | 0.077 | 0.081 | 0.084 | 0.253 | 0.256 | 0.260 | 0.237 | 0.241 | 0.241 |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where the outcome variable (score or self-concept), is regressed against TSR, the predictor of interest. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and cluster standard errors at school level.

**Naïve Estimates of Relationship Between TSR and Student Learning**

For comparison (to both existing literature and quasi-experimental evidence in the following) purposes, Tables 5 summarizes the relationship between TSR and student academic outcomes estimated by ordinary least squares (OLS) regression that accounts for school-fixed effects and clustering of standard errors. Consistent with the existing correlational literature, across all three subjects, TSR was significantly associated with increased academic outcomes, only the effect sizes for score outcomes (0.054, 0.089, and 0.049 SD in Chinese, English, and math, respectively) were much smaller than those in the literature (e.g., Cornelius-White, 2007), which is likely due to the improvement in my estimation precision (in particular, adding cubic functions of wave 1 scores in same and other subjects as well as cognitive test). On the other hand, TSR seemed to be more strongly related to student self-concept (0.202, 0.194, 0.158 SD in Chinese, English, and math, respectively) than to score and the effects sizes can be considered medium to large. These findings persisted when I estimated the model three times with each time adding one more set of student-, homeroom peer- and teacher-level covariates. To be clear, these OLS estimates are confounded by theoretically countless omitted variables and should not be interpreted as causal.

**IV Estimates of TSR Effects on Academic Outcomes**

Based on my identification strategy, the statistically adjusted estimates obtained from the value-added, 2SLS regression approach can be interpret as the causal effects of TSR on student learning. Table 6 reports the effects of TSR on student outcomes estimated by 2SLS regression that accounts for school-fixed effects and clustering standard errors at school level. I found that after correction for bias, one SD increase in TSR improved students' Chinese self-concept by approximately one SD (95% CI: 0.468-1.578) whereas on Chinese score, the estimate was educationally meaningful (0.202 SD, 95% CI: -0.300-0.704) but very imprecise. In English classrooms, TSR positively affected both score and self-concept: one SD increase in TSR improved student score by 0.665 SD (95% CI: 0.022-1.308) and self-concept by 0.855 SD (95% CI: 0.153-1.557). These findings were robust across three model specifications that add different sets of student-, homeroom peer-, and teacher-level covariates from wave 1.

These finding held true across a rich sets of robustness checks (Appendix Table A4, A5, and A6) using the same 2SLS regression: in Table A4, I add disruptive behavior or attention issue fixed effects; in Table A5, I add off-school tutoring fixed effects; in Table A6, I use a composite measure of TSR from principal component analysis (Appendix Table A1) as alternative predictor variable. Additionally, in Table A7, I use wave 2 CEPS cognitive test score as alternative outcome variable to show evidence that my 2SLS approach did not have overidentification issues.

**Discussions and Policy Implications**

Teacher is the most important measured aspect of schools in determining student achievement (Hanushek, 2011) and understanding more about the mechanisms through which teachers bring meaningful change to student outcomes is of central interests of researchers, policymakers, and educators. My study adds new, quasi-experimental evidence of the positive effects of teacher-student relationships on student academic outcomes. Using random assignment to a teacher-advisor as an instrument in an IV estimation, I found that one standard deviation (SD) increase in TSR because of the teacher-advisor significantly improved students' English score by 0.67 SD and self-concept in English and Chinese by 0.86 SD and 1.02 SD. The estimated effect of TSR on Chinese score (0.20 SD) was also educationally meaningful but I could not distinguish it from zero due the relatively large imprecision in the estimates. In all, the

Table 2.6. 2SLS estimates of TSR effects on student academic outcomes

| | Subject Score | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | 0.246<br>(0.248) | 0.195<br>(0.262) | 0.202<br>(0.256) | 0.768<br>(0.480) | 0.729*<br>(0.347) | 0.665*<br>(0.328) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.859** (df = 1; 62) | 7.066** (df = 1; 62) | 7.461** (df = 1; 62) | 3.259 (df = 1; 62) | 8.449** (df = 1; 62) | 9.233** (df = 1; 62) |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 |

| | Subject Self-Concept | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | 1.100**<br>(0.403) | 0.999**<br>(0.300) | 1.023***<br>(0.283) | 1.231<br>(0.680) | 1.144**<br>(0.417) | 0.855*<br>(0.358) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.859** (df = 1; 62) | 7.066** (df = 1; 62) | 7.461** (df = 1; 62) | 3.259 (df = 1; 62) | 8.449** (df = 1; 62) | 9.233** (df = 1; 62) |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate 2SLS regression where the outcome variable, score or self-concept, is regressed against the exogenously identified (by the instrumental variable) TSR. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects. Standard errors are clustered at school level.

fact that language teachers can consistently improve students' academic performance by using more praising, asking questions, and attention to individual students (three constructs used as the proxy for TSR) has promising implications to researchers, policymakers, and educators.

School advisory programs are evidence-based school reform efforts but the implementations require significant inputs from various parties including local government, school, teachers, and students, which leads to the critical interests of policymakers and educators in learning about whether advisors contribute meaningfully to student school outcomes. The current study adds novel evidence of teacher-advisor effects: being taught by teacher-advisor significantly improved student English and math score both by 0.14 SD and self-concept in Chinese and English by 0.22 SD and 0.21SD, with substantial effect sizes compared to the magnitude of teacher effects in value-added literature, e.g., Hanushek and Rivkin (2010) reviewed ten rigorous value-added studies leveraging within-school estimation and found that one SD increase in teacher quality improves student reading and math score by 0.13 and 0.17 SD. Broadly taken, on average, adding an additional advisor role to a content teacher has the approximately equivalent effects as improving teacher quality by one SD. Taking together the auxiliary nature of advisor effects, i.e., advisor effects were generated through social-emotional channels that parallel instruction, my findings have substantial policy implications for school leaders and educators who seek to redefine teacher's role in students' school life and find additional ways to impact students' learning beyond traditional classrooms.

There are at least four critical suggestions to appropriately interpret the findings of this study. First, I was not able to estimate TSR effects in math subject area but my findings do not suggest in any way that math teachers do not have causal effects on students. Since only the variation responding to the instrumental variable was used in the estimation and in math classrooms, TSRs did not differ between teacher-advisors and non-advisor teachers, the IV estimation did not produce any meaningful results for math teachers. Second, different IVs may lead to different estimates of effects because the IV estimates are "localized" around the instrument (only the variation responding to IV is used in the estimation). If future studies can identify other valid instruments, the results may differ. Third, in terms of external validity, the estimated TSR effects exist under the school setting that advisors and non-advisors teach same subjects, which is common in some countries (such as China, Japan, South Korea, and Israel) but not worldwide.

Lastly, I also note that teacher effects in this paper are identified as the within-school, between-teacher variance components of teachers' value added to student outcomes over a school year. Based on the random teacher-student assignment, this quasi-experimental design effectively accounts for the sorting of individual students to teachers, however, it does not rule out the possibility that teachers can influence individual students through peer effects. For example, a student learned more with a teacher not only because her teacher was effective in teaching and relationship-building but also her homeroom cohorts were all making progress with this teacher. Although this issue is mitigated by the various homeroom-level covariates I include in my estimation, it is helpful to bear in mind that the teacher effect in my study is broadly defined – it is blend of teacher's direct effect and indirect effect (through homeroom peers) on individual students.

In terms of limitations of this study,  I emphasize that a certain degree of external validity has been sacrificed during sample restriction process, where I only include 63 schools in my analytic samples to reenforce my internal validity. This led to systematic differences between schools included and excluded in my analyses: my sample schools appear to be more likely located in economically developed and urban areas and have significantly smaller class sizes. As a result, I note that my findings are not generalizable to schools in disadvantaged areas where schools are more likely to fail to restrictively implement random assignments. Further study should look closer to the assignment of teachers and students in schools from remote areas and if possible, conduct researcher-designed experiments to test the robustness of my findings.

I also note that due to the self-report nature of CEPS data, there are potentially large measurement errors embedded in the key variables. For instance, the self-concept measure only captures students' response to a single survey item therefore is potentially not accurately capturing the latent construct. Same applies to the TSR measure, which is proxied by three survey items rather than assessed using well-developed TSR batteries and questionnaires in the literature. Future research should be conducted after refining these measures.

In summary, the current study adds quasi-experimental evidence to the long-standing research question whether the observed relations between TSRs and student outcomes are causal and highlight that the social-emotional learning environments contribute meaningfully to student performance. These findings have substantial implications, on one hand, to researchers who aspire to understand more about the underlying mechanisms through which teachers affect students, and on the other hand, to policymakers and educators who seek evidence-based policy initiatives and educational practices to improve teacher effectiveness.

## APPENDIX. ADDITIONAL TABLES AND FIGURES

Table A1. Principal Components Analysis (PCA) on classroom TSR (based on full CEPS data)

Panel A. Chinese sample

| | Chinese classroom TSR | | |
|---|---|---|---|
| | Eigenvalue | Proportion of variance | Cumulative proportion |
| Comp1 | 1.4630 | 0.7130 | 0.7130 |
| Comp2 | 0.7203 | 0.1729 | 0.8860 |
| Comp3 | 0.5848 | 0.1140 | 1.0000 |

| | Principal components (eigenvectors) | | |
|---|---|---|---|
| | Comp1 | Comp2 | Comp3 |
| praise | 0.5516 | 0.7931 | -0.2583 |
| question | 0.6029 | -0.1651 | 0.7806 |
| attention | 0.5765 | -0.5863 | -0.5692 |

Panel B. English sample

| | English classroom TSR | | |
|---|---|---|---|
| | Eigenvalue | Proportion of variance | Cumulative proportion |
| Comp1 | 1.4528 | 0.7036 | 0.7036 |
| Comp2 | 0.7402 | 0.1827 | 0.8862 |
| Comp3 | 0.5842 | 0.1138 | 1.0000 |

| | Principal components (eigenvectors) | | |
|---|---|---|---|
| | Comp1 | Comp2 | Comp3 |
| praise | 0.5436 | 0.8109 | -0.2167 |
| question | 0.6063 | -0.2008 | 0.7695 |
| attention | 0.5805 | -0.5496 | -0.6008 |

Panel C. Math sample

| | Math classroom TSR | | |
|---|---|---|---|
| | Eigenvalue | Proportion of variance | Cumulative proportion |
| Comp1 | 1.4534 | 0.7041 | 0.7041 |
| Comp2 | 0.7383 | 0.1817 | 0.8858 |
| Comp3 | 0.5853 | 0.1142 | 1.0000 |

| | Principal components (eigenvectors) | | |
|---|---|---|---|
| | Comp1 | Comp2 | Comp3 |
| praise | 0.5453 | 0.8051 | 0.2335 |
| question | 0.6062 | -0.1863 | -0.7732 |
| attention | 0.5790 | -0.5631 | 0.5897 |

Notes: The survey questions are: “In your Chinese/English/math class, to what extent do you agree (0 = strongly disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = strongly agree) with the following statements”:
“My teacher always praises me” (praise)
“My teacher always asks me to answer questions” (question)
“My teacher always pays attention to me” (attention)

Table A2. Covariates balance check between teacher-advisors and non-advisor teachers

| | Teacher-Advisor | Non-Advisor Teacher | *p*-value |
|---|---|---|---|
| | N = 104 | N = 251 | |
| Female | 74.04% | 76.10% | 0.7 |
| Age | 39 (7) | 39 (7) | >0.9 |
| Experience (yrs) | 16 (8) | 16 (8) | >0.9 |
| Highest degree | | | 0.9 |
| Associate | 8.65% | 7.97% | |
| Bachelor | 88.46% | 87.65% | |
| Graduate | 2.88% | 4.38% | |
| Subject area | | | 0.7 |
| Chinese | 30.77% | 34.26% | |
| English | 32.69% | 33.07% | |
| Math | 36.54% | 32.67% | |

Notes: Cells report mean and standard deviation for continuous variables and percentage of each category for categorical variables. The *p*-statistic was obtained from a) Pearson's Chi-squared test for gender, degree, and subject area, and b) Wilcoxon rank sum test for teacher age and experience.

Table A3. Placebo test: teacher-advisor effects on cognitive outcomes

| | Wave 2 CEPS Cognitive Test Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English | Math | Math | Math |
| Taught by advisor | -0.055<br>(0.047) | -0.056<br>(0.041) | -0.065<br>(0.040) | -0.053<br>(0.040) | -0.055<br>(0.038) | -0.065<br>(0.038) | 0.047<br>(0.039) | 0.057<br>(0.039) | 0.058<br>(0.035) |
| Student Covariates | X | X | X | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X | | X | X |
| Teacher Covariates | | | X | | | X | | | X |
| School FE | X | X | X | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X | X | X | X |
| Observations | 5013 | 5013 | 5013 | 5039 | 5039 | 5039 | 5063 | 5063 | 5063 |
| R2 | 0.580 | 0.584 | 0.586 | 0.579 | 0.582 | 0.583 | 0.579 | 0.582 | 0.584 |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate OLS regression where the alternative outcome variable, wave 2 cognitive score, is regressed against the indicator of being taught by teacher-advisor. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and cluster standard errors at school level.

Table A4. 2SLS robustness: adding student at risk fixed effects

| | Subject Score | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | 0.246 | 0.194 | 0.202 | 0.772 | 0.728* | 0.660* |
| | (0.250) | (0.266) | (0.259) | (0.488) | (0.347) | (0.321) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| Risk-factor FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.711** (df = 1; 62) | 6.978** (df = 1; 62) | 7.556** (df = 1; 62) | 3.206 (df = 1; 62) | 8.746** (df = 1; 62) | 9.866** (df = 1; 62) |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 |

| | Subject Self-Concept | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | 1.098** | 1.002** | 1.027*** | 1.237 | 1.140** | 0.855* |
| | (0.406) | (0.303) | (0.285) | (0.685) | (0.414) | (0.352) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| Risk-factor FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.711** (df = 1; 62) | 6.978** (df = 1; 62) | 7.556** (df = 1; 62) | 3.206 (df = 1; 62) | 8.746** (df = 1; 62) | 9.866** (df = 1; 62) |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate 2SLS regression where the outcome variable, score or self-concept, is regressed against the exogenously identified TSR (by the instrumental variable, being taught by teacher-advisor). Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and student risk-factor fixed effects. Standard errors are clustered at school level.

Table A5. 2SLS robustness: adding student off-school tutoring fixed effects

| | Subject Score | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | 0.248<br>(0.253) | 0.194<br>(0.267) | 0.201<br>(0.260) | 0.782<br>(0.498) | 0.735*<br>(0.349) | 0.662*<br>(0.320) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| OS Tutoring FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.603**<br>(df = 1; 62) | 6.783**<br>(df = 1; 62) | 7.117**<br>(df = 1; 62) | 3.219<br>(df = 1; 62) | 8.761**<br>(df = 1; 62) | 10.066**<br>(df = 1; 62) |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 |

| | Subject Self-Concept | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | 1.133**<br>(0.420) | 1.033**<br>(0.315) | 1.058***<br>(0.297) | 1.247<br>(0.699) | 1.146**<br>(0.418) | 0.849*<br>(0.348) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| OS Tutoring FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.603**<br>(df = 1; 62) | 6.783**<br>(df = 1; 62) | 7.117**<br>(df = 1; 62) | 3.219<br>(df = 1; 62) | 8.761**<br>(df = 1; 62) | 10.066**<br>(df = 1; 62) |
| Observations | 5055 | 5055 | 5055 | 5080 | 5080 | 5080 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate 2SLS regression where the outcome variable, score or self-concept, is regressed against the exogenously identified TSR (by the instrumental variable, being taught by teacher-advisor). Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and off-school tutoring fixed effects. Standard errors are clustered at school level.

Table A6. 2SLS robustness: using alternative TSR measure (from PCA) as predictor

| | Subject Score | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR (PCA) | 0.179<br>(0.175) | 0.137<br>(0.183) | 0.143<br>(0.180) | 0.524<br>(0.319) | 0.491*<br>(0.222) | 0.466*<br>(0.221) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 8.028**<br>(df = 1; 62) | 7.465**<br>(df = 1; 62) | 7.922**<br>(df = 1; 62) | 3.405<br>(df = 1; 62) | 9.21**<br>(df = 1; 62) | 9.791**<br>(df = 1; 62) |
| Observations | 5009 | 5009 | 5009 | 5049 | 5049 | 5049 |

| | Subject Self-Concept | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR (PCA) | 0.759**<br>(0.274) | 0.687**<br>(0.200) | 0.713***<br>(0.191) | 0.849<br>(0.436) | 0.774**<br>(0.260) | 0.602*<br>(0.236) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 8.028**<br>(df = 1; 62) | 7.465**<br>(df = 1; 62) | 7.922**<br>(df = 1; 62) | 3.405<br>(df = 1; 62) | 9.21**<br>(df = 1; 62) | 9.791**<br>(df = 1; 62) |
| Observations | 5009 | 5009 | 5009 | 5049 | 5049 | 5049 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate 2SLS regression where the outcome variable, score or self-concept, is regressed against the exogenously identified (by the instrumental variable) TSR measure from PCA. Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects. Standard errors are clustered at school level.

Table A7. Placebo test: 2SLS estimates of TSR effects on cognitive score

| | Wave 2 CEPS Cognitive Test Score | | | | | |
|---|---|---|---|---|---|---|
| | Chinese | Chinese | Chinese | English | English | English |
| TSR | -0.274<br>(0.258) | -0.260<br>(0.222) | -0.309<br>(0.222) | -0.384<br>(0.400) | -0.283<br>(0.236) | -0.299<br>(0.212) |
| Student Covariates | X | X | X | X | X | X |
| Homeroom Covariates | | X | X | | X | X |
| Teacher Covariates | | | X | | | X |
| School FE | X | X | X | X | X | X |
| School clustered SE | X | X | X | X | X | X |
| 1st stage F-Statistics | 7.764**<br>(df = 1; 62) | 6.971**<br>(df = 1; 62) | 7.276**<br>(df = 1; 62) | 3.641<br>(df = 1; 62) | 9.792**<br>(df = 1; 62) | 10.76**<br>(df = 1; 62) |
| Observations | 5013 | 5013 | 5013 | 5039 | 5039 | 5039 |

*Notes*: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Cells report coefficients and associated standard errors in parentheses. Each column reports results of a separate 2SLS regression where the alternative outcome variable, wave 2 cognitive test score, is regressed against the exogenously identified TSR (by the instrumental variable, being taught by teacher-advisor). Each sample is estimated three times: the first model accounts for the cubic functions of the student's four wave 1 scores (in Chinese, English, math, and CEPS cognitive test) as well as wave 1 demographic characteristics including gender, age, single-child status, rural residency, migrant-worker family status, mother and father's total years of education, and family wealth; the second model adds homeroom size and the means of the wave 1 demographics of the student's homeroom peers, and the third model further adds teacher gender, age, education attainment, and experience. All models control for school fixed effects and student risk-factor fixed effects. Standard errors are clustered at school level.

## REFERENCES CITED


Angrist, Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association, 91(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Angrist, & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. The Journal of Economic Perspectives, 15(4), 69–85. https://doi.org/10.1257/jep.15.4.69

Aspy, & Roebuck, F. N. (1982). Affective Education: Sound Investment. Educational Leadership, 39(7), 488.

Aspy, D., Roebuck, Flora N., & National Consortium for Humanizing Education. (1977). Kids don't learn from people they don't like. Amherst, Mass.: Human Resource Development Press.

Baiocchi, Cheng, J., & Small, D. S. (2014). Instrumental variable methods for causal inference. Statistics in Medicine, 33(13), 2297–2340. https://doi.org/10.1002/sim.6128

Becker, & Luthar, S. S. (2002). Social-Emotional Factors Affecting Achievement Outcomes Among Disadvantaged Students: Closing the Achievement Gap. Educational Psychologist, 37(4), 197–214. https://doi.org/10.1207/S15326985EP3704_1

Birch, & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. Journal of School Psychology, 35(1), 61–79. https://doi.org/10.1016/S0022-4405(96)00029-5

Bound, Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. Journal of the American Statistical Association, 90(430), 443–450. https://doi.org/10.1080/01621459.1995.10476536

Bowlby, J. (1969). Attachment and loss: Vol. 1. Attachment. New York, NY: Basic Books.

Carrell, & Hoekstra, M. (2014). Are school counselors an effective education input? Economics Letters, 125(1), 66–69. https://doi.org/10.1016/j.econlet.2014.07.020

Chen, & Zhao, C. (2022). More is less: Homeroom teachers' administrative duties and students' achievements in China. Teaching and Teacher Education, 119, 103857–. https://doi.org/10.1016/j.tate.2022.103857

Christensen, C. M. (1960). Relationship between pupil achievement, pupil affect-need, teacher warmth, and teacher permissiveness. Journal of Educational Psychology, 51(3), 169–174. https://doi.org/10.1037/h0044666

Cornelius-White, J. (2007). Learner-Centered Teacher-Student Relationships Are Effective: A Meta-Analysis. Review of Educational Research, 77(1), 113–143. https://doi.org/10.3102/003465430298563

Crosnoe, Johnson, M. K., & Elder, G. H. (2004). Intergenerational bonding in school: The behavioral and contextual correlates of student-teacher relationships. Sociology of Education, 77(1), 60–81. https://doi.org/10.1177/003804070407700103

Davis. (2003). Conceptualizing the Role and Influence of Student-Teacher Relationships on Children's Social and Cognitive Development. Educational Psychologist, 38(4), 207–234. https://doi.org/10.1207/S15326985EP3804_2

Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T, Haynes, N. M., et al. (1997). Promoting social and emotional learning: Guidelines for educators. Alexandria, VA: Association for Supervision and Curriculum Development.

Fan, & Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. Educational Psychology Review, 13(1), 1–22. https://doi.org/10.1023/A:1009048817385

Galassi, Gulledge, S. A., & Cox, N. D. (1997). Middle School Advisories: Retrospect and Prospect. Review of Educational Research, 67(3), 301–338. https://doi.org/10.3102/00346543067003301

Garrett, & Steinberg, M. P. (2015). Examining Teacher Effectiveness Using Classroom Observation Scores: Evidence From the Randomization of Teachers to Students. Educational Evaluation and Policy Analysis, 37(2), 224–242. https://doi.org/10.3102/0162373714537551

Hanushek. (2011). The economic value of higher teacher quality. Economics of Education Review, 30(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

Hanushek, & Rivkin, S. G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. The American Economic Review, 100(2), 267–271. https://doi.org/10.1257/aer.100.2.267

Ho. (2003). Students' Self-Esteem in an Asian Educational System: The Contribution of Parental Involvement and Parental Investment. The School Community Journal, 13(1), 65–.

Hoffman. (2009). Reflecting on Social Emotional Learning: A Critical Perspective on Trends in the United States. Review of Educational Research, 79(2), 533–556. https://doi.org/10.3102/0034654308325184

Hurwitz, & Howell, J. (2014). Estimating Causal Impacts of School Counselors With Regression Discontinuity Designs. Journal of Counseling and Development, 92(3), 316–327. https://doi.org/10.1002/j.1556-6676.2014.00159.x

Kleibergen, & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics, 133(1), 97–126. https://doi.org/10.1016/j.jeconom.2005.02.011

Lee. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. International Journal of Educational Research, 53, 330–340. https://doi.org/10.1016/j.ijer.2012.04.006

Moore, Maggin, D. M., Thompson, K. M., Gordon, J. R., Daniels, S., & Lang, L. E. (2019). Evidence Review for Teacher Praise to Improve Students' Classroom Behavior. Journal of Positive Behavior Interventions, 21(1), 3–18. https://doi.org/10.1177/1098300718766657

McClure, Yonezawa, S., & Jones, M. (2010). Can school structures improve teacher-student relationships? The relationship between advisory programs, personalization and students' academic achievement. Education Policy Analysis Archives, 18, 17–. https://doi.org/10.14507/epaa.v18n17.2010

Mulhern, C. (2020). Beyond teachers: Estimating individual guidance counselors' effects on educational attainment. Unpublished Manuscript, RAND Corporation.

Murray, & Malmgren, K. (2005). Implementing a teacher–student relationship program in a high-poverty urban school: Effects on social, emotional, and academic adjustment and lessons learned. Journal of School Psychology, 43(2), 137–152. https://doi.org/10.1016/j.jsp.2005.01.003

Pianta, R. C. (1999). Enhancing relationships between children and teachers. American Psychological Association.

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), Handbook of research on student engagement (pp. 365–386). Springer Science.

Quin. (2017). Longitudinal and Contextual Associations Between Teacher-Student Relationships and Student Engagement: A Systematic Review. Review of Educational Research, 87(2), 345–387. https://doi.org/10.3102/0034654316669434

Rhodes, Spencer, R., Keller, T. E., Liang, B., & Noam, G. (2006). A model for the influence of mentoring relationships on youth development. Journal of Community Psychology, 34(6), 691–707. https://doi.org/10.1002/jcop.20124

Roorda, Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The Influence of Affective Teacher-Student Relationships on Students' School Engagement and Achievement: A

Meta-Analytic Approach. Review of Educational Research, 81(4), 493–529. https://doi.org/10.3102/0034654311421793

Staiger, & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. Econometrica, 65(3), 557–586. https://doi.org/10.2307/2171753

Wang, & Yang, D. (2021). Do homeroom teachers affect students' academic achievement in China's middle schools? Applied Economics Letters, 28(4), 329–333. https://doi.org/10.1080/13504851.2020.1752897

Xu, Zhang, Q., & Zhou, X. (2022). The Impact of Low-Ability Peers on Cognitive and Noncognitive Outcomes: Random Assignment Evidence on the Effects and Operating Channels. The Journal of Human Resources, 57(2), 555–596. https://doi.org/10.3368/jhr.57.2.0718-9637R2

Yang, Zubizarreta, J. R., Small, D. S., Lorch, S., & Rosenbaum, P. R. (2014). Dissonant Conclusions When Testing the Validity of an Instrumental Variable. The American Statistician, 68(4), 253–263. https://doi.org/10.1080/00031305.2014.962764

# CHAPTER IV. CONCLUSION

The overarching goal of this dissertation is to understand more about what are the teacher-related factors that drive student academic outcomes. Leveraging the national trend of random teacher-student assignment in China and analyzing a nationally representative, longitudinal, student-level data, I examine two sets of teacher-related factors and estimate whether they bring meaningful change to student learning: teacher human capital characteristics frequently used in human resources decisions and classroom teacher-student relationships.

Overall, my findings highlight the fact that the mechanisms through which teachers affect student outcomes are multidimensional, nuanced, and varying across different subject areas. In Chinese language arts, teachers' major in Education, graduate-level degree, and relationship with students all add meaningfully to students' learning, with more reliable effects on students' self-concept rather than score outcomes. In English as foreign language,  teachers' human capital profile does not matter but positive teacher-student relationships consistently improves both score and self-concept outcomes. In math, however, I did not find the similar theme and how math teachers affect student learning remains a challenging question.

My findings on Chinese junior high school students and their teachers also provide a comparison to the current literature that is largely focused on the Western, especially the U.S. education context. An important consensus between the two sides of the world is that teacher experience is not a determinator of teacher quality and a sound, evidence-based human resources policy metrics should not emphasize teaching experience. Another theme is that classroom teacher-student relationships and the social-emotional learning environment matter. Advisor programs are grounded in theory and integrating advising and mentoring responsibilities into teachers' role may be an effective policy initiative to improve school outcomes.