INTERNATIONAL ECONOMIC REVIEW
Vol. 65, No. 1, February 2024 DOI: 10.1111/iere.12668

TUTORING EFFICACY, HOUSEHOLD SUBSTITUTION, AND STUDENT ACHIEVEMENT: EXPERIMENTAL EVIDENCE FROM AN AFTER-SCHOOL TUTORING PROGRAM IN RURAL CHINA*

By Jere R. Behrman, C. Simon Fan, Naijia Guo, Xiangdong Wei, Hongliang Zhang, and Junsen Zhang

University of Pennsylvania, USA; Lingnan University, Hong Kong; The University of Hong Kong, Hong Kong; Zhejiang University, China

After-school tutoring has risen globally despite limited evidence of effectiveness. We implement a randomized after-school tutoring program in rural China where many children are left-behind by parents in care of grandparents. Compared to tutees cared for by parents, those in care of grandparents reported much smaller home-tutoring reductions but larger test-score gains. We interpret our data analysis with a model with tutoring efficacy and substitution between private and public inputs both differing by family background: Increased public tutoring generates larger test-score gains for children who experience greater tutoring efficacy and lesser substitution with household inputs, consistent with our estimates.

1. INTRODUCTION

Almost everywhere in the world, children's access to after-school learning opportunities, in the form of either direct family involvement or supplementary educational services, is highly dependent on family background (e.g., Weiss et al., 2009; Bray and Lykins, 2012). For example, in the United States, top-quintile families spend seven times more than bottom-quintile families on child-enrichment activities (Duncan and Murnane, 2016) and higher-educated parents also allocate much more time to educational child-care activities (Guryan et al., 2008; Ramey and Ramey, 2010). A strong positive relationship between family socioeconomic status (SES) and student participation in private supplementary education has also been documented in other countries, such as China (Zhang and Xie, 2016), South Korea (Kim and Lee, 2010), Japan (Matsuoka, 2015), and Poland (Safarzyńska, 2013). Partly owing to concern over the implications of after-school learning opportunities on child development and educational inequality, public provision of after-school learning support has risen globally. In the United States, the 21st Century Community Learning Centers (21st CCLC) program was authorized

*Manuscript received January 2022; revised April 2023.

We thank Joshua Angrist, Jonathan Davis, Weili Ding, Esther Duflo, Steven Lehrer, Leign Linden, Xiang Ma, Mark Rosenzweig, and seminar participants at the National University of Singapore, Peking University, Shanghai Jiao Tong University, Seoul National University, 2016 Asian Family in Transition Conference, and 2019 Asian and Australasian Society of Labour Economists Conference for valuable discussion and comments. We are also grateful to the Educational Bureau of the Longhui County in China for administrative support, and Chuantao Cui, Sudong Hua, Bei Luo, Amy Ru Chien Tseng, Baojun Wang, Kunlun Wang, Jia Wu, and Xiang Zhou for excellent research assistance. We acknowledge funding support provided by the National Natural Science Foundation of China (No. 72034006, 71173178), Hong Kong Research Grants Council General Research Fund (No. 458910), and the Chinese University of Hong Kong Research Committee Group Research Scheme. This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0005294. The study obtained approval from the Chinese University of Hong Kong Survey and Behavioural Research Ethics Committee and informed consent from the guardians of the students involved in the experiment. Please address correspondence to: Junsen Zhang, School of Economics, Zhejiang University, Hangzhou, China 310058. E-mail: *jszhang@cuhk.edu.hk*

[Correction added on 22 December 2023, after first Online publication: Funding information was updated.]

under the No Child Left Behind (NCLB) Act as an after-school program model to provide academic enrichment and services during nonschool hours to help students attending high-poverty, underperforming schools meet federal and state standards in core academic subjects.[1] In China, the State Council launched a national campaign in 2018 tightening scrutiny of its over 400,000 private after-school institutions[2] and strengthening the public role in providing after-school services (State Council of China, 2018).

In part due to the enormous policy interest, there have been a number of recent studies on the effects of after-school programs on children's academic outcomes. Nickow et al. (2020) provide an excellent review and meta-analysis of 96 experimental studies of the impacts of one-on-one or small-group tutoring on student learning in pre-K through 12. They elect to restrict their meta-analysis to tutoring offered by teachers, paraprofessionals, nonprofessionals, and parents, and find these programs "yield consistent and substantial positive impacts on learning outcomes" across a wide range of program characteristics and contexts, with effect sizes averaging over a third of a standard deviation.[3]

The review by Nickow et al. (2020) also raises two important gaps in the literature. First, although over 90% of school-age children live in low- and middle-income countries (LMICs), the studies summarized are primarily on high-income countries (HICs); for example, though the majority of the study titles refer to HICs, only two of the 96 study titles refer to LMICs (Banerjee et al., 2015; Conn, 2017). Second, these studies estimate the gross or total policy effects of tutoring programs that combine the ceteris paribus effects holding other inputs constant and the indirect effects through changes in private inputs after household reoptimization (Todd and Wolpin, 2003, 2007). Although some differences in the estimates are undoubtedly due to differences in program features, differences in household behavioral responses in providing private inputs also may be important sources of variation of the estimated gross policy effects (Behrman et al., 2013) but are nonetheless largely overlooked in prior research in this area.

In this article, we address these gaps. We first present reduced-form estimates based on an experiment that we conducted with randomly assigned after-school tutoring in rural China in which high-achieving fourth and fifth graders provided high-dosage, one-on-one tutoring to low-achieving second and third graders. We take advantage of the widespread "left-behind children" (LBC) phenomenon in rural China where many children are living apart from both parents who have migrated to work in cities by implementing the experiment in a rural Chinese county with a high prevalence of LBC to assess household behavioral responses and treatment–effect heterogeneities between children with and without parents at home. Prior to the experiment, LBC—most of whom were cared for by grandparents—received far less home tutoring compared to non-LBC. During the tutoring program, non-LBC tutees reported large and significant reductions in home tutoring at both the extensive and intensive margins, whereas LBC tutees had much smaller, and often insignificant, reductions in home tutoring. We find that the tutoring program significantly improved tutees' endline math scores and the score gains were significantly larger for LBC. However, we find little evidence that

[1] The 21[st] CCLC program was first authorized by Congress in 1994 as a community-learning-center model to open up public schools for broader use to provide academic enrichment and recreational activities after school for all community members (James-Burdumy et al., 2007). In 2002, the NCLB reauthorized the 21[st] CCLC program and narrowed it to focus only on academic content to complement in-school learning. In the 2013–14 school academic year, total federal funding of over $1.1 billion was allocated to implement the 21[st] CCLC program serving approximately 1.7 million students (U.S. Department of Education, 2015). In addition, school districts were also mandated to use a portion of their Title I funding to offer supplementary educational services to students attending schools that have failed their "adequate yearly progress" goals for three consecutive years.

[2] This number is from the following China Daily article:http://www.chinadaily.com.cn/a/201811/22/WS5bf60804a310eff30328a54e.html(last accessed April 18, 2019).

[3] They also find that on average, effects are larger for teacher and paraprofessional tutoring programs than for nonprofessional and parent tutoring and that effects tend to be larger for the earlier grades. Most of these studies are on at-risk children due to family background or other factors and some of the studies find larger effects within their samples for more disadvantaged children.

the program increased tutees' endline reading scores. The disparity in the achievement effects between math and reading is consistent with some previous work on the efficacy of other educational interventions across subjects that also tend to find these interventions more effective for improving math scores than reading scores (e.g., Black et al., 2008; Angrist et al., 2010; Abdulkadiroğlu et al., 2011; Dobbie and Fryer, 2013; Muralidharan et al., 2019), but contrasts with Nickow et al.'s (2020) overall summary of similar effects for reading and math. However, teacher and paraprofessional tutoring account for the majority of the tutoring programs reviewed in Nickow et al. (2020), and they note that disaggregating by subject and tutor type yields much smaller effect sizes for nonprofessional reading tutoring programs by volunteers, the category most similar to our peer-tutoring program.

In order to examine possible mechanisms, we next develop a simple structural model of student achievement production in which caregivers (i.e., parents or grandparents) maximize their valuation of child achievement net of the total cost of home-tutoring provision subject to a value-added achievement production function. The distributions of valuation of child achievement differ between parents and grandparents, permitting parents on average to value child achievement more than grandparents. The total costs of home tutoring include fixed costs, which is consistent with corner solutions of no home tutoring for some families, in addition to marginal costs related to home-tutoring time. The value-added production function, inter alia, includes total tutoring, which is a constant elasticity of substitution (CES) function of tutoring time received in school and tutoring time received at home. This function allows (i) differential efficacy parameters between children cared for by grandparents versus those cared for by parents, (ii) diminishing returns to aggregate total tutoring received after school, and (iii) complementarity between school-based, public-tutoring inputs and home-based, private-tutoring inputs in generating total tutoring.

If school tutoring and home tutoring are direct substitutes in promoting achievement as our estimates below suggest, we show that several propositions follow from this model: (i) Households' optimal home-tutoring time decreases with school-tutoring time; (ii) the positive direct effect of increased school-tutoring time dominates the negative indirect effect through optimal substitution for home-tutoring time so that both total tutoring and student achievement increase; (iii) given the same level of school-tutoring time, children cared for by caregivers with higher returns to total tutoring (driven by higher valuation of achievement or greater tutoring efficacy) receive more home tutoring; (iv) if school-tutoring time increases, caregivers with higher returns to total tutoring reduce their home-tutoring time more; and (v) if school-tutoring time increases, children with caregivers with higher return to total tutoring experience smaller increases in total tutoring.

We then proceed to estimate the model using data of the randomized after-school tutoring experiment described above. We find that in the CES function for total tutoring, diminishing returns dominate complementarity so school tutoring and home tutoring are direct substitutes. We also find that although tutoring is more effective for more-disadvantaged children in the sense of being LBC, parents value children's achievement much more than grandparents. The greater valuation of children's achievement by parents than by grandparents dominates the lower tutoring efficacy for non-LBC than LBC, leading to parents on average devoting more time to home tutoring than grandparents.

Using the estimated baseline model, we first decompose the difference in the predicted treatment effect between LBC and non-LBC into the difference in their home-tutoring time and the difference in their tutoring efficacy. We find that the time substitution accounts for one-eighth of the difference and the difference in tutoring efficacy accounts for seven-eighths of the difference. We also conduct counterfactual experiments by setting different levels of school tutoring for LBC and non-LBC and evaluating the effects on their home tutoring and test scores. We find that as we increase school-tutoring time, the marginal benefits decline for both LBC and non-LBC but that the benefits are always larger for LBC at any given level of school-tutoring time, implying that the optimal allocation of school-tutoring time would generally favor LBC than non-LBC.

This article makes several contributions to the literature on the effects of public-school inputs on private household inputs and student achievement. First, by integrating public and private inputs to produce student achievement through two competing mechanisms (i.e., diminishing returns and complementarity), our model provides a framework that encompasses the divergent evidence of household responses to public-school inputs found in different settings and contexts in prior research. For example, whereas Gelber and Isen (2013) show that parental involvement with children increases in response to access to the Head Start Program in the United States, studies of developing contexts generally find evidence for the substitution between household inputs and school resources (e.g., Pop-Eleches and Urquiola, 2013; Yuan and Zhang, 2015; Mbiti et al., 2019). Also in the developing contexts of India and Zambia, Das et al. (2013) find that household behavioral responses play significant roles in determining the overall effectiveness of increases in school resources: Only unanticipated increases in school resources that are not offset by reductions in households' own educational spending have positive effects on test scores, whereas anticipated increases in school resources that crowd-out private household educational expenditures generate no test score gains.

Second, we extend the empirical investigation of this strand of literature to increases in public inputs in after-school learning support and exploit the "left-behind children" phenomenon in rural China to examine heterogeneity in household behavioral responses and treatment effects. By linking the differences in the extent of substitution of home-tutoring inputs and in the efficacy of those inputs between children with and without parents at home to the difference in their test score gains, we are able to gauge more directly the implications of household behavioral responses on the total effect of the intervention.

Third, because the extent of household responses in home-tutoring inputs is quantitatively small and statistically insignificant for children living apart from their parents, our estimated policy effect on test scores for children left-behind by both parents places a sharper upper bound on the direct production function effect (i.e., the ceteris paribus effect) compared to other contexts with more room for household substitution of private inputs.

Since only students who had scored below the class median in the baseline test were eligible to be tutees in the after-school tutoring program, the empirical analysis of the article is also related to the literature on remedial educational interventions targeted to children lagging behind academically (e.g., Banerjee et al., 2007, 2010; Duflo et al., 2011), particularly those employing randomized controlled trials (RCTs) to evaluate the effectiveness of high-dosage, small-group, or one-on-one tutoring outside the regular school hours (e.g. Cabezas et al., 2011; Cook et al., 2014; Weiss et al., 2019).[4] Our article contributes to this strand of research in three respects. First, we demonstrate that remedial educational interventions are most effective for children who come from disadvantaged family backgrounds. Second, we highlight both theoretically and empirically the important roles of household substitution behaviors in determining the effectiveness of remedial educational programs. Third, our results also imply that remedial educational programs offering learning inputs in settings that are subject to lesser substitution of household inputs and/or the inputs are more efficacious will generate larger effects.

Finally, this article also adds to an emerging literature demonstrating that children from disadvantaged background benefit the most from public educational interventions, such as universal child-care programs (e.g., Cascio and Schanzenbach, 2013; Bitler et al., 2014; Havnes and Mogstad, 2015) and charter schools or injecting successful charter school practices into traditional public schools (e.g., Angrist et al., 2012; Fryer, 2014). For example, Kottelenberg and Lehrer (2017) and Cornelissen et al. (2018) find that in both Quebec and Germany,

[4] It is also worth mentioning a peer-to-peer learning intervention conducted by Li et al. (2014) in a Beijing school that enrolls exclusively children of migrant workers. The intervention combines the treatments of group incentives and peer interactions—the latter of which may also include peer tutoring outside regular school hours—by pairing high- and low-achieving classmates as benchmates and offering the pairs group incentives for the lower-achiever's score improvement. They find that the intervention increases lower achievers' test scores without affecting the high achievers.

children from disadvantaged family background benefit the most from universal child-care programs, and Angrist et al. (2013) show that Massachusetts charter schools adhering to the "No Excuse" model[5] are most effective for poor non-Whites and low-baseline achievers. Our findings of larger score gains for children living apart from both parents suggest that focusing the limited public resources for after-school learning support on children with disadvantaged family backgrounds not only provides a cost-effective means of improving the learning of children lagging behind academically but also acts as an important equalizer for reducing inequality in child development by family background.

We organize our study as follows. Section 2 provides background on the educational system and "left-behind children" phenomenon in China. Section 3 develops a model to guide our analysis. Section 4 describes our after-school tutoring experiment. Section 5 presents our results on student achievement. Section 6 discusses our results on home-tutoring inputs. Section 7 concludes.

2. BACKGROUND

According to the first large-scale national survey on household educational expenditure carried out in 2017, over one-third of elementary and secondary school students in China participated in academically focused after-school programs and spent on average RMB 5,021 (or USD 733) per pupil per year (Wei, 2019). However, there is a large urban–rural disparity in after-school program expenditures with urban households on average spending 3.6 times that spent by their rural counterparts (Wei, 2019).[6] Accompanying this difference in shadow education use is an even more striking urban–rural disparity in children's educational attainments. Using a nationwide data set of college entrance examinations and admissions, Li et al. (2015) show that rural youth from poor counties are eight times less likely (2% vs. 16%) than urban youth to attend a four-year college.[7] Differences in access to after-school learning opportunities, combined with differences in school qualities, have turned the nation's educational system, once a great equalizer, into an inequality exaggerator. Partly owing to such concerns, in 2018, the State Council launched a national campaign tightening the regulation of private after-school institutions and demanding public elementary and secondary schools to enhance their roles in offering after-school services. In order to wrest control from the frenzied private-tutoring industry, local governments have used public budgets to provide varied after-school learning support, including running optional extra-hour after-school programs, offering public-school teachers overtime pay to provide one-on-one tutoring through online platforms, and providing free access to services purchased from private-service providers.

Another important feature of the educational system in rural China is the unprecedented scale of children left-behind by both parents. According to the 2010 Population Census, over 61 million children aged 17 years or below were living without one or both parents, of which 46.7% were left-behind by both parents (All-China Women's Federation, 2013). That is, one in ten children in China—or one in six children in rural China—is living apart from both parents, a scale unprecedented elsewhere in the world. In a previous study, we find significant negative impacts of being left-behind by both parents, but not by one parent, on rural Chinese children's cognitive development (Zhang et al., 2014). Moreover, we also find that the absence of both parents is associated with significantly lower household educational inputs in the form of family direct involvement (e.g., homework checking, home tutoring), whereas the absence of one parent is not, suggesting that the negative achievement effects of being left-behind by

[5] The "No Excuse" pedagogy emphasizes discipline, traditional reading and math skills, selective teacher hiring, and increased instructional time.

[6] In a case study of Chongqing, Zhang (2014) finds that students' participation rates in private tutoring are 74% for urban schools, 34% for county schools, but only 11% for township/village schools.

[7] Zhang et al. (2015) and Zhao et al. (2017) also show that rural children score significantly lower than their urban counterparts in math, vocabulary, and cognitive ability tests using the Chinese Family Panel Survey (CFPS) and the China Education Panel Survey (CEPS), respectively.

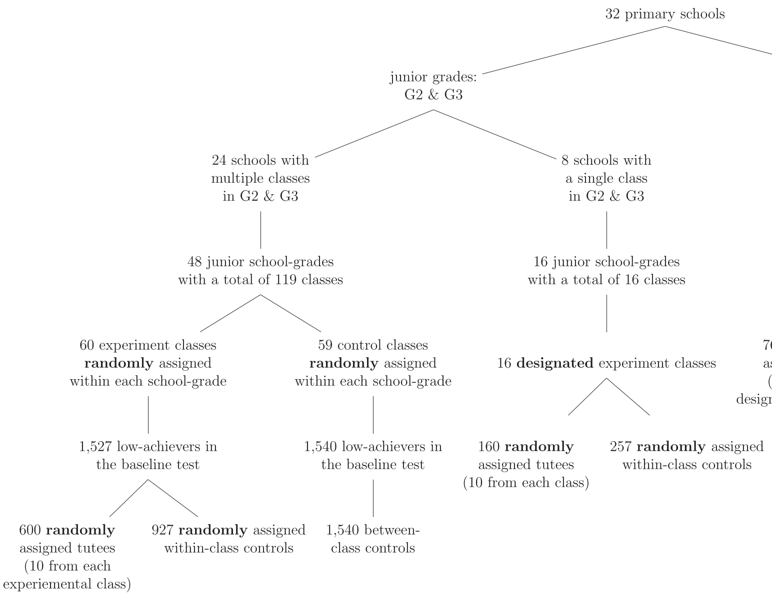


FIGURE 1

EXPERIMENT DESIGN CHART

both parents may work (at least in part) through the lack of after-school learning support at home. Given the important "dynamic complementarities" between early life learning outcomes and later-life human-capital investments (e.g., Cunha and Heckman, 2007; Heckman, 2007; Aizer and Cunha, 2012), left-behind children's disadvantage in early life learning caused by parental absence can lead to considerable losses in their lifetime human capital. Thus, if the disadvantage in after-school learning support is indeed a main mechanism through which parental absence impedes children's learning progress, public provision of after-school learning support may be an effective compensatory intervention for ameliorating the negative learning effects of absence of both parents.

3. AFTER-SCHOOL TUTORING EXPERIMENT

3.1. *Program Description and Random Assignment.* Our randomized after-school tutoring experiment was conducted in Longhui County of Hunan Province, a designated national poverty county with a high prevalence of left-behind children. In 2010, the county's per-capita GDP (RMB 7,000) was less than a quarter of the national average (RMB 30,000), 90% of its 1.2 million residents were rural, and more than one-third of its primary- and middle-school-age children were left behind by both parents.[8] We implemented the intervention as a supervised peer-tutoring program in which high-achieving fourth and fifth graders were paired, respectively, with low-achieving second and third graders of the same primary school to offer one-on-one tutoring after school hours with professional-teacher support.[9] We defined students as "high-achieving" (or "low-achieving") if their combined scores in math and reading in the baseline test were above (or below) the medians of their classes. Figure 1 illustrates the detailed process for the selection of tutees and tutors for the peer-tutoring program. With help from the county's educational bureau, we recruited 32 primary schools to participate in the peer-tutoring program, of which 24 schools had multiple classes in each grade and

[8] See Zhang et al. (2014) for a more detailed discussion of the background of this county.

[9] It is worth noting that this peer-tutoring program only targets to improve the outcomes of the tutees and we also find no effects of participating in the peer-tutoring program on tutors' test scores.

eight schools had only a single class in each grade. For the 24 schools with multiple classes in grade 2/3, we randomly assigned half of the classes in each grade (a total of 60 classes) as experiment classes—from which the tutees were selected—and the remaining half (a total of 59 classes) as control classes.[10] However, for the eight schools with only a single class in grade 2/3, all of the 16 classes were designated as experiment classes because all schools were assured the opportunity to participate in the program when they were recruited. Therefore, we ended up with a total of 76 experiment classes, of which 60 classes were randomly selected and had at least one control class from the same school grade.

For each experiment class, we randomly selected 10 low-achieving students (i.e., those with baseline scores below the class median) to participate in the program as tutees, whereas the remaining unselected low-achieving students were assigned as within-class controls. Our main analysis thus compares 760 tutees and 1,184 within-class controls randomly selected from the 76 experiment classes. As a robustness check for the potential existence of externalities of tutees' participation in the program on control students from the same classes, in Online Appendix A, we further compare the tutees from the subset of 60 experiment classes randomly selected from the school grades with multiple classes to all low-achieving students in the unselected control classes from the same school grades (i.e., between-class controls). The between-class estimates, reported in Online Appendix Table A1, show little evidence for the existence of within-class externalities.

For each grade 2/3 experimental class, the school assigned a grade 4/5 class for us to recruit tutors among the class's high achievers in the baseline test. In order to attract high achievers to participate in the program, the principals of all participating schools announced at the beginning of the school year that all participating tutors would be given certificates of merit after completion of their service. Thanks to this arrangement and the cooperation of the head teachers, for every targeted senior class, the number of high-achieving students who applied to be a tutor with the consent of their parents/guardians exceeded the quota needed and we randomly selected 10 tutors from the pool of oversubscribed applicants.[11]

The tutoring experiment lasted for eight months from November 2013 to June 2014, with a one-month winter break between mid-January and mid-February 2014. During the experimental period, the randomly paired tutors and tutees met in designated tutorial rooms (usually the tutees' classrooms) for 45-minute tutorials Mondays through Thursdays. Each tutorial room hosted 10 tutor–tutee pairs. A teacher was recruited from a grade other than the tutees' and tutors' grades to supervise each tutorial room. The teachers' duties included keeping attendance, maintaining discipline, and having weekly meetings with every tutor–tutee pair to review the progress and recommend individually customized target areas to improve the tutee's competencies. However, as the key feature of the intervention is to deliver *personalized* tutoring, the teachers were asked not to conduct any group teaching directly, although they still answered individual questions per request. Before the start of the intervention, the tutors had attended two training sessions offered by the educational professionals recruited by the research team. During the training, the tutors were told to focus their tutoring practices on three types of activities, all targeted to offer personalized instruction to students lagging behind in classroom learning. First, the tutor helped the tutee to review the teaching content in today's class that the tutee did not comprehend. Second, the tutor supervised the tutee to finish the selected assignments from homework that the tutee had difficulty in solving independently. Third, the tutor offered individually customized instruction in areas where the tutee was lacking competencies, which could be identified through either teachers' recommendations, tutees' self-reflections, or tutors' observations.

[10] The number of classes in each junior school grade with multiple classes ranged from two to four. When a school grade had three classes, a coin was first flipped to decide whether to select one or two experimental classes from this school grade.

[11] In results not reported in this article, we also compare the randomly selected tutors to the unselected eligible applicants in the endline test and find no evidence that participating in the program affected tutors' test scores.

During the intervention, we conducted two rounds of supplementary surveys (in the first week of December 2013 and May 2014, respectively) asking the tutors to report the allocations of tutoring time across different activities and subjects and the teachers to record attendance and subjective assessments (on a scale of 0 to 5) of the respective performances of individual tutors and tutees. The combined summary statistics of the two rounds of supplementary surveys are shown in Table A2 in the Online Appendix. The tutorial sessions were well attended with an average attendance rate of 91%. Offering individually customized instruction in areas where the tutees were lacking competencies accounted for the highest share of total tutoring time (36%), followed by supervising homework (30%) and reviewing contemporary learning content (27%). Tutors on average spent more time on math than reading (53% vs. 39%), with less than 10% spent on other subjects.[12] However, there was little difference between tutees living with and without parents in terms of attendance rates, time allocations, and teachers' subjective assessments of the efforts of tutees/tutors.

Although this experiment started with 760 assigned pairs, only 90% of these pairs lasted until the end. Although most of the terminations were caused by tutees or tutors switching schools, in some cases either tutees or tutors decided to withdraw from the experiment though still enrolled in the same school. Whenever a tutee switched school or simply withdrew from the experiment, we suspended the pair. However, if a tutor switched school or withdrew from the experiment, we replaced him/her with another tutor applicant not selected initially. Nonetheless, throughout all empirical analyses, only the initial tutee or tutor assignment status is used.

3.2. *Data.* We conducted two survey rounds: a baseline survey in October 2013 and an endline survey in June 2014. The baseline survey consisted of a student questionnaire asking each student's age, gender, time allocations after school (including home-tutoring time), and a household questionnaire asking information on family composition, parents' ages, schooling attainments, and migration status. When at least one parent stayed at home, the household questionnaire was filled out by a parent; otherwise, it was filled out by the primary caregiver, who was asked to verify the information by phone with the student's parents. In the latter case, information on the primary caregiver was also collected. The endline survey was conducted in June 2014, about two weeks before the end of the 2013–14 school year.

A baseline test on math and reading was conducted in September 2013, a month before the baseline survey. Students' cumulative scores in math and reading in this baseline test were used to determine their eligibility for participating in the after-school tutoring program in the role of tutees for second and third graders or tutors for fourth and fifth graders. At the end of June 2014, an endline test on math and reading was administered to evaluate the achievement effects of the program. Both rounds of the tests were centrally administered and graded by teachers from different schools assigned by the educational bureau. For the endline test, we also recruited teachers from different schools as enumerators to proctor the examinations.

Panel A of Table 1 checks the balance of the baseline individual characteristics between tutees and controls. Parental absence is indeed a pervasive phenomenon in this sample: Among 760 tutees, 44% had both parents absent from home and 30% had one parent absent from home, leaving only 26% living with both parents. From self-reported student surveys, only 30% of the tutees received tutoring at home before the experiment. Conditional on receiving tutoring at home, the average home-tutoring time was 3.9 hours per week, even higher than the dosage of our experiment (three hours per week). However, with the majority not receiving any home tutoring at all, the unconditional average home-tutoring time was only 1.2 hours per week. In all analyses of this article, individual test scores are converted to z-scores with respect to the distribution of scores in the control classes on the same test. In the baseline test, tutees on average scored 0.70σ and 0.69σ below the means of the control classes in

[12] Some (but not all) schools offered English classes to grades 2 and 3 students, in which case the majority of the tutoring time spent on other subjects was on English.

TABLE 1
SUMMARY STATISTICS, BALANCE CHECKS, AND SAMPLE SELECTION

	Tutees' Mean (1)	Tutees - Controls Difference (2)
Panel A. Sample Balance Checks		
Living w/ one parent only	0.301	0.012
	(0.459)	(0.021)
Living w/ no parent	0.437	−0.023
	(0.496)	(0.023)
Fraction receiving tutoring at home	0.303	−0.003
	(0.460)	(0.019)
Average weekly home-tutoring hours (excluding 0)	3.890	−0.166
	(2.523)	(0.231)
Average weekly home-tutoring hours (including 0)	1.177	−0.063
	(2.262)	(0.108)
Baseline math scores	−0.701	−0.026
	(0.946)	(0.037)
Baseline reading scores	−0.691	−0.012
	(0.984)	(0.037)
Panel B. Sample Selection		
Fraction taking the endline test	0.925	−.024
	(0.263)	(0.011)
Number of classes	76	76
Number of students	760	1944

NOTE: The sample consists of all low-achieving grades 2 and 3 students in all the 76 experiment classes who scored below the class medians in the baseline test, among which 10 from each class were randomly assigned as tutees and the remaining were assigned as within-class controls. For each baseline variable denoted by the row heading, column 1 reports the tutees' mean and column 2 reports the difference in the means between the tutees and controls after adjusting for class fixed effects. Reported in parentheses are standard deviations for means and standard errors for differences.

math and reading, respectively. Column 2 compares tutees and their within-class controls in the full sample of 76 experimental classes and finds no evidence of any significant differences in parental absence status, home-tutoring time, or baseline scores. Table A2 in the Online Appendix conducts separate balance checks for the subsamples of students living with and without parents and also finds no evidence of any significant differences in the baseline characteristics between tutees and their within-class controls within each subsample.

Panel B of Table 1 compares the test-taking status in the endline test between the tutees and controls. Among a total of 760 tutees, 703 (or 92.5%) took the endline test, which is 2.4 percentage points lower than control students from the same experimental classes. Although small in magnitude, this difference is statistically significant at the 5% level. Thus, we cannot completely rule out the possibility that participating in the after-school tutoring program may have resulted in some students not taking the endline test. In particular, if the head teachers of some experimental classes or the supervisors of some tutorial sessions had discouraged/restricted some tutees who had made poorer progress in the program than their peers from taking the endline test, our estimated achievement effects of the program would be biased upward. Given that the 2.4 percentage point difference in the test-taking rate is quantitatively quite small, we conduct our main empirical analysis by comparing 703 tutees and 1,112 within-class controls who had taken the endline test without considering the potential sample selection problem, but perform a partial identification analysis in Online Appendix B following a trimming procedure proposed by Lee (2009)[13] to construct conservative bounds

[13] See also Lucas and Mbiti (2012), Zhang (2016), and Mills and Wolf (2017) for applications of Lee's (2009) trimming method in estimating bounds on treatment effects on student achievement in educational settings.

TABLE 2
AVERAGE TREATMENT EFFECTS ON MATH AND READING SCORES

	Math Scores		Reading Scores	
	(1)	(2)	(3)	(4)
Tutee	0.134***	0.136***	0.007	0.017
	(0.040)	(0.040)	(0.040)	(0.040)
Baseline scores	0.543***	0.543***	0.685***	0.657***
	(0.043)	(0.044)	(0.043)	(0.045)
Baseline individual controls		Y		Y
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of observations	1815	1815	1815	1815

NOTE: The sample consists of all tutees and controls from 76 experiment classes whose endline test scores are available. Each column corresponds to a separate regression. The dependent variable is the test scores in the endline test (math for columns 1 and 2 and reading for columns 3 and 4). All columns control for class fixed effects. Columns 2 and 4 further control for baseline individual characteristics including gender and dummy indicators for living with only one parent and living with no parent in the baseline. Robust standard errors clustered at the class level are reported in parentheses. $***p < 0.01$.

on the treatment effects taking into account the presence of differential sample selection between tutees and controls.

4. EMPIRICAL RESULTS

4.1. *Average-Achievement Effect.* In this subsection, we use the experimental data from the randomized after-school tutoring program to conduct reduced-form estimations of the average-treatment effect on student achievement. Specifically, we estimate the following value-added achievement regression by comparing tutees and their within-class controls from the same experimental classes:

$$Y_{ij1} = \alpha Y_{ij0} + \lambda D_{ij} + \varphi_{j1} + \varepsilon_{ij1}, \tag{1}$$

where Y_{ij1} denotes the endline test scores of student i from class j, Y_{ij0} denotes the baseline test scores of student i, D_{ij} is a dummy indicator equal to 1 if student i was assigned to be a tutee and 0 if otherwise, φ_{j1} is a class fixed effect that captures the unobserved contemporary determinants of achievement progress shared in common among all students from class j, and ε_{ij1} is an error term, consisting of both an individual-level component and a class-level component. Two remarks are worth noting for the specification of Equation (1). First, the coefficient estimate on the lagged baseline test scores α is expected to be strictly between 0 and 1 because of the decay of past input effects and measurement error in test scores (Todd and Wolpin, 2003). Second, the coefficient on the treatment dummy λ identifies a "policy" effect—the average *total* effect of the tutoring program after taking into account household behavioral responses in home inputs—instead of the ceteris paribus effect holding home inputs constant.

In Table 2, columns 1 and 2 estimate the average-treatment effect on tutees' math scores using, respectively, the baseline specification Equation (1) and an expanded version of Equation (1) that further controls for baseline individual characteristics including gender and parental absence status. Estimates of the average-treatment effect coefficient λ change relatively little across specifications. Taking column 2 with the full set of control variables as an example, the estimated coefficient $\hat{\lambda}$ indicates that the tutees have an average gain of 0.136σ in math scores (significant at 1% level) compared to the control students from the same experimental classes. Columns 3 and 4 replicate the same estimations for tutees' reading scores.

Regardless of the empirical specifications used, the estimates of λ for reading scores are always insignificant and small in magnitude.

The disparity in the estimated achievement effects between math and reading, though somewhat striking, does not come as a complete surprise. First, anecdotal evidence collected from the teachers supervising the tutorial rooms indicates that the pairs usually spent more time working on math than reading during the tutorial sessions, which may reflect the difference in the nature of homework for these two subjects: math homework requires a lot of written work and calculations while language homework emphasizes more reading and reciting that do not need to be done in front of a tutor. Second, one may well expect that the value of tutoring may differ across subjects. For example, in mathematics, there are many problem-solving "tricks," which can be best learned through tutoring. For language, on the other hand, a student can comprehend new words only if he/she spends effort to memorize them. Thus, if tutoring increases test scores largely through improving problem-solving and test-taking skills, which are more important for math than reading, the achievement effects of tutoring would also be more salient for math than reading. Moreover, the fact that the estimated coefficient on lagged test scores is always greater for reading than math also suggests that the achievement gap in reading is more persistent over time and thus more difficult to close compared to that in math. Indeed, previous work on the efficacy of other educational interventions across subject areas generally find that achievement gains tend to be larger for math than reading (e.g., Angrist et al., 2010; Abdulkadiroğlu et al., 2011; Angrist et al., 2012; Dobbie and Fryer, 2013; Fryer, 2014). For example, in a similar setting of a randomized-controlled evaluation of 27 centers of the *enhanced* 21^{st} CCLC program in the United States with 45 minutes of each daily session used for structured-academic instruction similar to our after-school tutoring program, Black et al. (2008) also find significant positive effects on math scores, but no effects on reading scores. Because we never obtain any statistically significant or quantitatively large estimates for reading scores, for the more-detailed analysis in the remaining of the article, we focus only on the effects of the program on tutees' math scores.

4.2. *Differential Achievement Effects by Parental Absence Status.* As mentioned, an emerging literature demonstrates that children from disadvantaged backgrounds benefit the most from public-educational interventions. In this subsection, we use parental-absence status in the baseline to proxy for family background and investigate the heterogeneity in the achievement effects for tutees of different parental-absence statuses.

In order to allow the treatment effects to vary by tutees' parental-absence statuses, we first interact the treatment dummy (D_{ij}) with three mutually exclusive dummy indicators representing a student's parental-absence status in the baseline: namely, $NoParent_{ij}$ for whether a student was living without both parents, $OneParent_{ij}$ for whether a student was living with a single parent, and $TwoParents_{ij}$ for whether a student was living with both parents. Column 1 of Table 3 reports estimates of an expanded version of Equation (1) by replacing the treatment dummy (D_{ij}) with the three interaction terms. The point estimates show that the average treatment effect is 0.091σ for tutees living with both parents, 0.076σ for tutees living with a single parent, and 0.203σ for tutees living without both parents. Although the first two estimates are not significant at conventional levels, the last coefficient is significant at the 1% level. The inclusion of baseline individual controls in column 2 has little effect on the estimates of these coefficients. Since the estimated coefficients for tutees living with one and both parents are very close to each other in both specifications, we pool these two categories together and use a single dummy indicator $Parent_{ij}$ to denote whether a student was living with at least one parent in the baseline. Columns 3 and 4 report estimates employing only two interaction terms $D_{ij} \times NoParent_{ij}$ and $D_{ij} \times Parent_{ij}$. The estimates in column 3 suggest that the program improved the math scores by 0.083σ for tutees living with one or both parents and by 0.203σ for those living without both parents. Whereas the latter estimate remains significant at the 1% level, the former also becomes significant at the 10% level because of the increased precision in the estimation after pooling students living with one and both parents

TABLE 3
DIFFERENTIAL TREATMENT EFFECTS BY PARENTAL ABSENCE STATUS

	(1)	(2)	(3)	(4)
Tutee∗Living w/ both parents	0.091	0.107		
	(0.068)	(0.086)		
Tutee∗Living w/ one parent only	0.076	0.084		
	(0.052)	(0.056)		
Tutee∗Living with one or both parents			0.083*	0.094*
			(0.047)	(0.053)
Tutee*Living w/ no parent	0.203***	0.190***	0.203***	0.190***
	(0.046)	(0.049)	(0.046)	(0.049)
Baseline individual controls		Y		Y
Baseline scores	Y	Y	Y	Y
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of observations	1815	1815	1815	1815

NOTE: The sample consists of all tutees and controls from 76 experiment classes whose endline test scores are available. Each column corresponds to a separate regression. The dependent variable the math scores in the endline test. All columns control for baseline scores and class fixed effects. Columns 2 and 4 further control for baseline individual characteristics including gender and dummy indicators for living with only one parent and living with no parent in the baseline. Robust standard errors clustered at the class level are reported in parentheses. ∗∗∗$p<0.01$, ∗$p<0.1$.

into a combined category. Moreover, the difference in these point estimates (0.120σ, with a p-value of 0.02) indicates that the program yielded larger achievement gains for tutees cared for by grandparents than those cared for directly by parents. Overall, the results in Table 3 point to larger math achievement gains for tutees living without both parents, suggesting an important role of parental-absence status in affecting the efficacy of the after-school tutoring intervention implemented.

4.3. *Accounting for Other Potential Sources of Treatment Heterogeneity.* One may be concerned about whether parental-absence status is indeed the most relevant characteristic to define subgroups with heterogeneous treatment effects given that parental absence is correlated with other student and family characteristics, which could also be potential sources of treatment heterogeneity. In order to investigate this possibility, we estimate alternative empirical specifications that allow the treatment effects to differ by four other exogenous/predetermined characteristics, namely baseline scores, gender, whether having received any tutoring at home in the baseline, and the weekly home-tutoring time in the baseline. The results of these regressions are reported in Panel A of Table 4. For all the characteristics examined, the coefficients on their interaction terms with the treatment dummy are very small quantitatively and statistically insignificant. Moreover, the coefficient estimates on the treatment dummy itself, which range from 0.122σ to 0.151σ, remain statistically significant and quantitatively similar to the estimate (0.136σ) obtained from the specification in column 3 of Table 2 without any interaction term.

In Panel B of Table 4, we further investigate whether the difference in the estimated achievement effects between tutees living with parent(s) and those living without both parents is robust to accounting for other potential sources of treatment heterogeneity. The inclusion of the additional interaction term between the treatment dummy and the corresponding characteristic examined in each column of Panel B of Table 4 has relatively little effect on the estimated coefficients on $D_{ij} \times NoParent_{ij}$ and $D_{ij} \times Parent_{ij}$.

4.4. *Treatment Heterogeneity with the Causal Forest.* In Subsections 4.2 and 4.3, we examine treatment heterogeneity by choosing several one-way interactions between the treatment dummy and exogenous/predetermined student characteristics. True variability in treatment effects, however, may work through more complicated mechanisms and depend on flexible,

TABLE 4
ROBUST ANALYSIS OF OTHER POTENTIAL SOURCES OF TREATMENT HETEROGENEITY

	(1)	(2)	(3)	(4)
Panel A. Other dimensions of treatment heterogeneity				
Tutee	0.122**	0.123**	0.143***	0.151***
	(0.048)	(0.048)	(0.040)	(0.039)
Tutee*baseline scores	−0.021			
	(0.047)			
Tutee*female		0.034		
		(0.075)		
Tutee*tutored at home in the baseline			−0.022	
			(0.071)	
Tutee*home-tutoring time in the baseline				−0.011
				(0.018)
Baseline individual controls	Y	Y	Y	Y
Baseline scores	Y	Y	Y	Y
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of observations	1815	1815	1815	1815
Panel B. Robustness of differential treatment effects by parental absence status				
Tutee*Living with one or both parents	0.084	0.084	0.094*	0.107*
	(0.060)	(0.059)	(0.053)	(0.049)
Tutee*Living with no parent	0.177***	0.178***	0.196***	0.201***
	(0.056)	(0.058)	(0.050)	(0.049)
Tutee*baseline scores	−0.017			
	(0.047)			
Tutee*female		0.029		
		(0.075)		
Tutee*tutored at home in the baseline			−0.007	
			(0.070)	
Tutee*home-tutoring time in the baseline				−0.008
				(0.018)
Baseline individual controls	Y	Y	Y	Y
Baseline scores	Y	Y	Y	Y
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of observations	1815	1815	1815	1815

NOTE: The sample consists of tutees and their within-class controls in all the 76 experimental classes. The dependent variable is the math scores in the endline test. Each column in each panel corresponds to a separate regression. All regressions control for baseline scores, gender, dummy indicators for living with only one parent and living with no parent in the baseline, and class fixed effects. The regressions in columns 3 and 4 further control for a dummy indicator for being tutored at home in the baseline and the weekly home-tutoring hours in the baseline, respectively. Robust standard errors clustered at the class level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

multiway interactions. In this subsection, we apply the causal-forest method[14] proposed by Wager and Athey (2018) to predict treatment effects based on high-dimensional, nonlinear functions of observed variables.

Following Davis and Heller (2020), who carefully document the implementation details of their application of causal forest to estimate treatment heterogeneity of a Chicago summer-jobs program, we adopt the following estimation steps in our context:

[14] The causal-forest method shares many similar principles as the conventional classification and regression trees (CART) and random forests but differs in two fundamental ways (Athey and Imbens, 2016). First, it aims to estimate conditional average-treatment effects (CATEs) instead of predicting outcomes. Second, it adopts an "honest" approach to estimation, whereby the training sample is split into two parts, one for constructing the tree (including the cross-validation step) and the other for estimating treatment effects within leaves of the tree.

(i) Draw 2/3 observations of the data without replacement as a training sample.
(ii) Use a random half of this training sample to grow a tree via recursively partitioning a set of covariates[15] that maximizes the variance of treatment effects across the leaves according to the cross-validation criterion. An important assumption for the causal forest is that within each leaf, treatment assignment is orthogonal to potential outcomes. For this assumption to hold under stratified randomization within each class in our context, we employ the sample weight option of the *causal_forest* package in R using inverse-probability weights that equal to $1/\bar{D}_j$ for treatment observations and $1/(1-\bar{D}_j)$ for control observations, where $\bar{D}_j$ is the probability of treatment of class j.
(iii) Switch to the remaining half of the training sample, assign the new observations to the corresponding leaves of the tree constructed in step (ii) based on their covariates, and calculate the estimated treatment effect for each leaf as the difference in the *weighted* mean outcomes between the treatment and control observations in this subset of the training sample assigned to that leaf with the weights defined in step (ii).
(iv) Repeat steps (i) to (iii) 30,000 times. Given that each training sample consists of 2/3 observations only, each observation is on average used in estimating 20,000 out of a total of 30,000 causal trees.
(v) For each observation, use the subset of the causal trees (i.e., on average 10,000) whose training sample does not contain this observation to make *out-of-bag* predictions of the treatment effect and average across these predictions to calculate a single-valued predicted-treatment effect.[16]

Figure 2 gives a visual presentation of how the predicted-treatment effects from the causal forest and actual-treatment effects are related. In this figure, we sort observations into 20 bins by percentile ranking of predicted-treatment effect (*x-axis*), and for each bin calculate the actual-treatment effect as the difference in mean outcomes between the treatment and control observations within that bin (*y-axis*). Both the scatter plots and the fitted-linear line demonstrate fairly strong positive relationships between the actual- and predicted-treatment effects, suggesting that treatment effects indeed differ across subgroups determined by high-dimensional, nonlinear functions of covariates as used in the causal-forest predictions.

In order to investigate whether the positive relationship between the actual- and predicted-treatment effects demonstrated in Figure 2 is statistically meaningful, we conduct three empirical tests. First, we follow Davis and Heller (2020) to estimate the following regression:

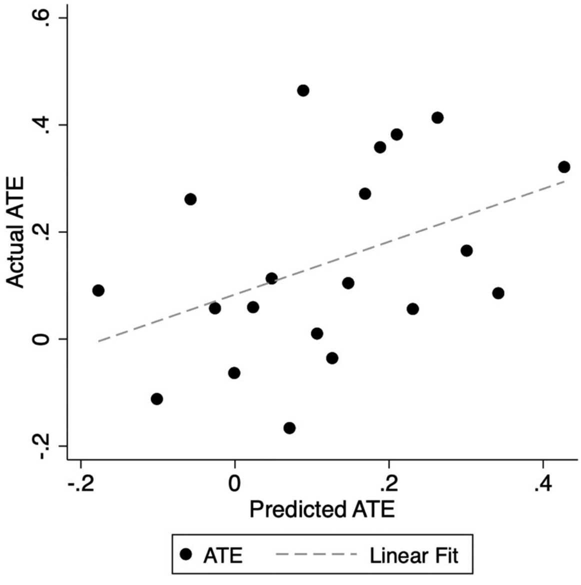
$$Y_{ij1} = \alpha + X_{ij}\beta + \lambda D_{ij} + \mu\hat{\tau}_{ij} + \eta D_{ij}\hat{\tau}_{ij} + \epsilon, \tag{2}$$

where X_{ij} is a vector including the randomization blocks (i.e., classes) and the covariates used in the causal-forest prediction and $\hat{\tau}_{ij}$ is the causal-forest prediction of the treatment effect. Rejecting $\eta = 0$ means both (a) there exists treatment heterogeneity and (b) that heterogeneity is captured by the causal-forest predictions based on the covariates.[17] Second, we estimate another functional form that Chernozhukov et al. (2020) show to yield the best-linear predic-

[15] The set of covariates used includes baseline scores, baseline home-tutoring time, and dummy indicators for female, living without both parents, and whether having received any tutoring at home in the baseline, all of which are employed in examining treatment heterogeneity through their interactions with the treatment dummy in the regressions in Tables 3 and 4.

[16] By averaging predicted-treatment effects across only the *out-of-bag* predictions made when an observation is used in neither determining the tree structure nor estimating the treatment effects within leaves, the predicted-treatment effect from this causal-forest procedure is not prone to overfitting.

[17] However, failing to reject $\eta = 0$ does not necessarily mean treatment effects are homogeneous. Even if treatment effects are heterogeneous, η could still be 0 if the causal-forest predictions do not "work" (i.e., contain too much noise relative to signal for predicting treatment heterogeneity).



NOTES: For each of the 20 bins defined by percentile of predicted treatment effects by the causal forest, the figure plots the actual average treatment effect against the predicted average treatment effect for each bin. The dashed line is the linear fitted line between the actual and predicted average treatment effects.

FIGURE 2

PREDICTED VS. ACTUAL TREATMENT EFFECTS

tor of the CATE using the causal-forest predicted treatment effect. In our context, their strategy corresponds to estimating the following linear regression:

$$Y_{ij1} = \alpha + \delta \hat{B}_{ij} + \lambda(D_{ij} - \bar{D}_j) + \eta(D_{ij} - \bar{D}_j)(\hat{\tau}_{ij} - \bar{\tau}) + \epsilon \tag{3}$$

weighted by

$$w_{ij} = \frac{1}{\bar{D}_j(1 - \bar{D}_j)},$$

where $\hat{B}_{ij}$ is the predicted-test score *without treatment* for observation i based on the same set of covariates X as in the causal forest,[18] $\bar{D}_j$ is the probability of treatment of class j as defined above, and $\bar{\tau}$ is the average-predicted-treatment effects for all observations from the causal forest. Third, we combine the first two strategies and estimate a hybrid specification that replaces $\hat{B}_{ij}$ in Equation (3) by the set of covariates X_{ij} used in Equation (2), that is,

$$Y_{ij1} = \alpha + \beta X_{ij} + \lambda(D_{ij} - \bar{D}_j) + \eta(D_{ij} - \bar{D}_j)(\hat{\tau}_{ij} - \bar{\tau}) + \epsilon. \tag{4}$$

Panel A of Table 5 presents the results of these three empirical tests. The estimates of the interactive coefficient ($\hat{\eta}$s) are positive and significant in all specifications, suggesting that variation in predicted-treatment effects from the causal forest indeed represents statistically meaningful heterogeneity in actual-treatment effects. However, the uncertainties of the estimation of the predicted-treatment effect $\hat{\tau}_{ij}$ (a right-side variable in all specifications) out of a random procedure are *not* accounted for in the standard errors of these estimated coefficients. In order to address this problem, we perform two robustness checks on the statistical

[18] Specifically, we estimate a random forest of 10,000 trees using the control observations only and average across their predictions to obtain $\hat{B}_{ij}$.

TABLE 5
ASSESSING THE CAUSAL FOREST PREDICTIONS

	(1)	(2)	(3)	
Panel A. Slope Test				
$D_{ij}\hat{\tau}_{ij}$	0.481**			
	(0.238)			
$(D_{ij} - \bar{D}_j)(\hat{\tau}_{ij} - \bar{\tau})$		0.686***	0.508**	
		(0.222)	(0.244)	
Predicted ΔY without treatment		Y		
Covariates	Y		Y	
Class fixed effects	Y		Y	
90% CI	[0.075, 0.856]	[0.305 1.031]	[0.091, 0.889]	
p-value, one-tailed permutation test	0.014	0.008	0.016	
	Tercile of predicted treatment effect			Difference
	T1	T2	T3	T3−T1
Panel B. Summary Statistics by Tercile of Predicted Treatment Effect				
Actual treatment effect	0.048	0.123**	0.239***	0.191**
	(0.062)	(0.062)	(0.061)	(0.087)
Tutee	0.372	0.385	0.405	0.033
				(0.028)
Female	0.365	0.379	0.412	0.046*
				(0.280)
Baseline scores	−0.566	−0.660	−0.694	−0.128**
				(0.053)
Living with no parent	0.370	0.430	0.521	0.150***
				(0.028)
Tutored at home in the baseline	0.365	0.233	0.357	−0.008
				(0.027)
Home-tutoring time in the baseline	1.595	0.891	1.400	−0.195
				(0.141)

NOTE: Panel A shows estimates of η for Equations (2)–(4) as discussed in Subsection 4.4. Panel B shows the means of the actual treatment effect and baseline characteristics for each tercile of predicted treatment effect from the causal forest (columns 1–3) and the difference between the first and third terciles (column 4). Robust standard errors are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

inference of the coefficient estimate $\hat{\eta}$. The first adopts a version of the bootstrap strategy proposed by Chernozhukov et al. (2020) to construct a valid confidence interval (CI) for $\hat{\eta}$ that accounts for variability in the causal-forest predicted treatment effect $\hat{\tau}_{ij}$, whereas the second calculates a one-tailed p-value for where our estimated coefficient $\hat{\eta}$ falls in the distribution of permuted estimates under the null hypothesis $\eta = 0$.[19] The implementation details of these two robustness checks are documented in the Online Appendix C.1. The results of these two robustness checks are reported at the bottom of Panel A of Table 5. For all empirical specifications, the lower bounds of the 90% CIs are positive and the one-tailed p-value of the permutation test is below 0.02.

Panel B of Table 5 reports descriptive statistics broken down by tercile of causal-forest predicted treatment effects. The first row presents another piece of evidence that the causal-forest prediction "works" for predicting actual-treatment heterogeneity: The top tercile group (T3) in terms of causal-forest predicted treatment effects also has the largest actual-average treatment effect (0.239σ), followed by the middle tercile group (T2, 0.123σ) and bottom tercile group (T1, 0.048σ), and the T3−T1 difference is statistically significant at the 5% level.

[19] For similar applications of this permutation strategy, see also Heckman et al. (2010), Campbell et al. (2014), and Davis and Heller (2020).

The rest of the panel tries to understand the possible mechanisms by comparing baseline student characteristics across different tercile groups. For every baseline characteristic, the sign of the T3−T1 difference is consistent with the sign of the coefficient estimate on its one-way interaction with the treatment dummy reported in Tables 3 or 4. However, only for gender, baseline scores, and parental absence, the T3−T1 difference is statistically significant. Specifically, compared with children predicted to be in the bottom tercile group, those predicted to be in the top tercile group are 4.6 percentage points (p-value < 0.1) more likely to be female, have average baseline scores 0.128σ lower (p-value < 0.05), and are 15.0 percentage points (p-value < 0.01) more likely to be living with no parent. Note that for each baseline characteristic, the T3−T1 difference only represents its simple correlation with the predicted-treatment effect, which may confound the true mechanisms through its correlation with other determinants of treatment heterogeneity. Therefore, we cannot take the statistically significant T3−T1 difference as decisive evidence that a baseline characteristic is indeed relevant in defining treatment heterogeneity.[20] Nonetheless, for the dummy indicator for living with no parent, the facts that the T3−T1 difference is largest both quantitatively and statistically among all baseline characteristics and that the coefficient estimate on its interaction term with the treatment dummy is also significant (Table 3) seem to suggest that it is indeed the most relevant characteristic to define subgroups with heterogeneous treatment effects.

4.5. *Results on Home-Tutoring Inputs.* We first examine the difference in baseline home-tutoring inputs by parental-absence status before the intervention (Subsection 4.5.1) and then compare changes in home-tutoring inputs after the intervention by students' treatment and parental-absence status (Subsection 4.5.2).

4.5.1. *Baseline home-tutoring inputs.* In this subsection, we examine whether and to what extent the amount of tutoring that students received at home before our intervention differs by their parental-absence status. Specifically, we estimate the following regression:

$$H_{ij0} = \beta_0 NoParent_{ij} + \beta_1 OneParent_{ij} + \beta_2 TwoParents_{ij} + \mu_j + \eta_{ij}, \tag{5}$$

where H_{ij0} is a self-reported measure of home-tutoring inputs for student i from class j in the baseline, taking the form of either a dummy indicator for having received any tutoring at home or the total home-tutoring hours in the week prior to the baseline survey, the three dummy indicators $NoParent_{ij}$, $OneParent_{ij}$, and $TwoParents_{ij}$ are the same as defined in Subsection 4.2, and μ_j is a class fixed effect. In estimating Equation (5), we normalize the class fixed effects to sum to 0 across all students in the sample: That is, $\sum_{j=1}^{J}\sum_{i=1}^{n_j}\mu_j = 0$, where n_j denotes the number of students from class j and J is the total number of classes. With this normalization, the coefficients β_0, β_1, and β_2 on the three mutually exclusive dummy indicators $NoParent_{ij}$, $OneParent_{ij}$, and $TwoParents_{ij}$ can be directly interpreted as the regression-adjusted, group-specific means in H_{ij0}.

Columns 1 and 2 in Panel A of Table 6 report estimates of Equation (5) with H_{ij0} measured by the dummy indicator for having received any tutoring at home in the week prior to the baseline survey and the total weekly home-tutoring hours, respectively. For children living with one and both parents, the proportions who had received tutoring at home are almost exactly the same (35.1% and 35.2%), both of which are substantially higher than that of children living without both parents (27.6%). The average home-tutoring time is also very similar between children living with a single parent (1.47 hours/week) and both parents (1.51 hours/week), but is much smaller for those living without both parents (1.05 hours/week). Since the absence of a single parent has little effect on the amount of tutoring that children

[20] This is also part of the reason why for gender and baseline scores the coefficient estimates on their interaction terms with the treatment dummy are insignificant in Panel A of Table 4, although their differences between the T3 and T1 group are statistically significant in Panel B of Table 5.

TABLE 6
RESULTS ON HOME-TUTORING INPUTS

	Tutored at Home (1)	Weekly Home-Tutoring Hours (2)	Tutored at Home (3)	Weekly Home-Tutoring Hours (4)
Panel A. Self-Reported Home-Tutoring Inputs in the Baseline				
(A1) Living w/ both parents	0.351***	1.505***	-	-
	(0.018)	(0.101)		
(A2) Living w/ one parent only	0.352***	1.474***	-	-
	(0.017)	(0.095)		
(A3) Living w/ parent(s)	-	-	0.352***	1.489***
			(0.012)	(0.069)
(A4) Living w/ no parent	0.276***	1.049***	0.276***	1.049***
	(0.014)	(0.078)	(0.014)	(0.078)
Difference by the absence of	-	-	0.076***	0.440***
both parents (i.e., (A3) − (A4))			(0.018)	(0.104)
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of students	1815	1815	1815	1815

	Tutees		Controls	
	Δ Tutored at Home (1)	Δ Weekly Home-Tutoring Hours (2)	Δ Tutored at Home (3)	Δ Weekly Home-Tutoring Hours (4)
Panel B. Changes in Self-Reported Home-Tutoring Inputs				
(B1) Living w/ parent(s)	−0.153***	−0.464***	−0.168***	−0.668***
	(0.022)	(0.122)	(0.018)	(0.115)
(B2) Living w/ no parent	−0.048*	−0.098	−0.146***	−0.568***
	(0.020)	(0.127)		
Difference by the absence of	0.104***	0.366**	0.022	0.100
both parents (i.e., (B1) − (B2))	(0.033)	(0.186)	(0.027)	(0.171)
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of students	703	703	1112	1112

NOTE: Each column of each panel corresponds to a separate regression. In Panel A, the dependent variable is a dummy indicator for having reported any tutoring at home in the baseline survey for the odd columns and the reported total home-tutoring hours in the week prior to the baseline survey for the even columns. In Panel B, the dependent variable is the change in the dummy indicator for having reported any tutoring at home between the baseline and endline surveys for the odd columns and the change in the reported weekly home-tutoring hours between the baseline and endline surveys. All regressions control for class fixed effects, which are normalized to sum to 0 across all students in the sample used in the regression. Robust standard errors clustered at the class level are reported in parentheses. $^{***}p < 0.01$,$^{**} p < 0.05$,$^{*} p < 0.1$.

received at home, we pool children living with one and both parents together into a single dummy indicator $Parent_{ij}$ as defined before. The estimates, reported in columns 3 and 4 in Panel A of Table 6, show that children living with parent(s) were 7.6 percentage points (or 27.5%) more likely to have received tutoring at home and on average reported 0.44 hours (or 41.9%) more home-tutoring time than those living apart from both parents, and the differences in both the extensive and intensive margins are significant at the 1% level.

4.5.2. *Changes in home-tutoring time.* For both the baseline and endline surveys, the self-reported home-tutoring inputs (H_{ijt}) are the sum of the *actual* home-tutoring inputs (H^*_{ijt}) and a reporting error (e_{ijt}), that is, $H_{ijt} = H^*_{ijt} + e_{ijt}$ for $t \in \{0, 1\}$. In the baseline, no student was aware of the after-school tutoring program to be implemented. Thus, any reporting error (if it exists) should be balanced between the tutees and controls because of the random assignment. However, in the endline, the implementation of the after-school tutoring program may have altered students' incentives and reporting behaviors, especially for controls. Some families in the control group seemed to believe that the program had targeted students most disadvantaged in home-tutoring inputs and therefore may have incentives to underreport their home-tutoring times in the endline. To the extent that the reporting behaviors of the tutees and controls were affected differently in the endline, we cannot directly compare changes in self-reported home-tutoring times between the tutees and controls in a pooled class fixed-effect regression.

Instead, we partition the tutees and controls into two subsamples, and for each subsample estimate a separate class fixed-effect regression of the changes in self-reported home-tutoring inputs as follows:

$$\Delta H_{ij} = H_{ij1} - H_{ij0} = \pi_0^{\kappa} NoParent_{ij} + \pi_1^{\kappa} Parent_{ij} + \delta_j^{\kappa} + \xi_{ij}^{\kappa}, \quad \kappa \in \{T, C\}. \tag{6}$$

The superscript $\kappa \in \{T, C\}$ denotes whether the equation is estimated using the tutees' subsample (T) or the controls subsample (C), and the class fixed-effect δ_j^{κ} is now normalized to sum to 0 across all students in each subsample. Columns 1 and 2 in Panel B of Table 6 report estimates of Equation (6) for the tutees' subsample. For tutees living with parent(s), the estimates of the coefficient on $Parent_{ij}$ (π_1^T) indicate that they were 15.3 percentage points less likely to report having received tutoring at home in the endline (column 1) and on average reported 0.46 fewer hours of home-tutoring time per week compared to the baseline (column 2). However, for tutees living apart from both parents, the estimates of the coefficient on $NoParent_{ij}$ (π_0^T) indicate much smaller reductions in the reported home-tutoring incidence and time, although the extensive-margin reduction (−4.8 percentage points) remains marginally significant at the 10% level. Despite the potential effect of participating in the after-school tutoring program on the change in reporting error Δe_{ij}, as long as such effect is the same for tutees with different initial parental-absence status, that is, $E[\Delta e_{ij} | Parent_{ij} = 1, D_{ij} = 1, \delta_j^T] = E[\Delta e_{ij} | NoParent_{ij} = 1, D_{ij} = 1, \delta_j^T]$, then the difference in the two coefficients $\pi_1^T - \pi_0^T$ still identifies the differential response in the *actual* home-tutoring inputs (ΔH_{ij}^*) between tutees living with and without parents, that is, $\pi_1^T - \pi_0^T = E[\Delta H_{ij}^* | Parent_{ij} = 1, D_{ij} = 1, \delta_j^T] - E[\Delta H_{ij}^* | NoParent_{ij} = 1, \ D_{ij} = 1, \delta_j^T]$. For both the extensive and intensive measures of home-tutoring inputs, the difference $\hat{\pi}_1^T - \hat{\pi}_0^T$, reported at the bottom of Panel A Table 6, is always negative and statistically significant, suggesting that there is indeed a differential extent of substitution in the *actual* home-tutoring inputs between tutees living with and without parents.

Columns 3 and 4 in Panel B of Table 6 report estimates of Equation (6) for the subsample of control students who did not participate in the after-school tutoring program themselves but observed the participation of some of their classmates. For these students, the effect of the program on changes in their reported home-tutoring inputs is likely to be dominated by changes in their reporting behavior (Δe_{ij}) instead of changes in *actual* home-tutoring inputs (ΔH_{ij}^*). Moreover, if the effect of the program on change in reporting error of control students is the same by initial parental-absence status, that is, $E[\Delta e_{ij} | Parent_{ij} = 1, D_{ij} = 0, \delta_j^C] = E[\Delta e_{ij} | NoParent_{ij} = 1, D_{ij} = 0, \delta_j^C]$, then changes in the self-reported home-tutoring inputs (ΔH_{ij}) should also not differ significantly between control students living with and without parents. Indeed, although both groups of control students reported large and significant reductions in home-tutoring times, the difference in the extent of the reduction is rather small quantitatively: −16.8 versus −14.6 percentage points at the extensive margin and −0.67 versus −0.57 hours/week at the intensive margin. Moreover, the fact that we cannot reject the equality of the changes in the self-reported home-tutoring inputs between the two types of control students (i.e., $\pi_1^C = \pi_0^C$) suggests no evidence that the program yields differential effects on reporting behavior for control students with different initial parental-absence status. If the conclusion of no differential program effects on reporting behavior also carries over to the tutees, then the estimate of $\pi_1^T - \pi_0^T$ in columns 1 and 2 in Panel B of Table 6 indeed identifies the differential response in the home-tutoring inputs for tutees with different parental-absence status in the baseline.

Although our results indicate that tutees cared for directly by parent(s) had experienced larger reductions in home-tutoring inputs and smaller achievement gains than tutees cared for by grandparents, a question remains as to what extent the difference in the achievement gains between the two types of tutees is attributable to the difference in household substitution of home-tutoring inputs. In order to answer this question, we next build and estimate a structural model of student achievement that allows a quantitative assessment of the difference in achievement gains between tutees with different parental-absence statuses into the difference in the direct effect due to the difference in the efficacy of tutoring inputs and the difference in the indirect effect due to the difference in the household substitution.

5. MODEL

In this section, we develop a simple structural model to integrate home- and school-tutoring inputs and household decision making about home tutoring into the determination of student achievement.

We extend the value-added production function developed by Summers and Wolfe (1977), Boardman and Murnane (1979), Hanushek (1979), and Todd and Wolpin (2003)[21] by relating the endline achievement of student i in class j (Y_{ij1}) to baseline achievement (Y_{ij0}), contemporaneous classroom-level common influences (δ_{j1}), contemporaneous after-school tutoring inputs (T_{i1}), and an error term (ϵ_{ij1}) as follows:

$$Y_{ij1} = \alpha Y_{ij0} + A_\tau T_{i1} + \delta_{j1} + \epsilon_{ij1}. \tag{7}$$

Although the conventional value-added model assumes that lagged achievement Y_{ij0} is a sufficient statistic for all historical factors (e.g., innate ability, family circumstances, past investments) affecting contemporaneous achievement, the theory of dynamic human capital formation suggests that early life circumstances and subsequent investments may interact to determine later learning outcomes (e.g., Heckman, 2007; Cunha et al., 2010).[22] Thus, it is possible that early life environments, for which we use parental absence status as a proxy, may interact with contemporaneous after-school tutoring inputs T_{i1}. Therefore, we allow the efficacy of total tutoring in promoting achievement A_τ to differ by caregiver type $\tau \in \{p, g\}$ to account for the potential complementarity or substitutability between early life environments and contemporaneous tutoring inputs, with p representing the child is cared for by parents and g representing the child is cared for by grandparents. We assume ϵ_{ij1}, which consists of the impact of other omitted contemporary variables and measurement error on endline test scores, is orthogonal to the included inputs in Equation (7) and is individually and independently distributed (i.i.d.) drawn from a normal distribution $N(0, \sigma^2)$.[23]

[21] Several papers compare estimates of teacher effectiveness derived from value-added models to test-score gains derived from random assignment of students to teachers and find the models to be reliable (e.g., Kane and Staiger, 2008; Kane et al., 2013; Chetty et al., 2014).

[22] Although some influential research supports the "dynamic complementary" of human capital formation by showing that the effects of preschool interventions are larger for those with higher initial endowments (e.g. Aizer and Cunha, 2012; Heckman et al., 2013), a few recent studies suggest the opposite by documenting that the latter life investments can help remediate for disadvantage generated very early in life (Adhvaryu et al., forthcoming; Goff et al., forthcoming; Rossin-Slater and Wüst, 2020).

[23] Given the random assignment of school-tutoring time S_{i1}, this assumption requires home-tutoring time H_{i1} to be uncorrelated with the omitted contemporary inputs and measurement error in ϵ_{ij1} conditional on lagged achievement Y_{ij0} and class fixed effects δ_{j1}.

Total tutoring in the endline T_{i1} is an aggregation of the tutoring time received in school (S_{i1}) and tutoring time received at home (H_{i1}) in the endline via a CES function with a degree of scale ρ as follows:

$$T_{i1} = [S_{i1}^{\gamma} + \theta H_{i1}^{\gamma}]^{\frac{\rho}{\gamma}}, \tag{8}$$

where the factor-loading parameter of school-tutoring time is normalized to be 1, and $\theta > 0$ is the factor-loading parameter of home-tutoring time. In a robustness check presented in Subsection 6.2, we allow θ to vary by caregiver's type or education, but we find no evidence of heterogeneity in home-tutoring efficiency. $\gamma \in (-\infty, 1)$ with $\frac{1}{1-\gamma}$ corresponding to the elasticity of substitution between S_{i1} and H_{i1}. ρ is the degree of scale and we focus on diminishing returns to scale, that is, $\rho \in (0, 1)$.

Given the achievement-production function in Equation (7), taking the cross derivative of endline test score (Y) with respect to endline school and home tutoring (S and H) yields

$$\frac{\partial^2 Y}{\partial S \partial H} = A_\tau \rho(\rho - \gamma)\theta[S^\gamma + \theta H^\gamma]^{\frac{\rho}{\gamma}-2} S^{\gamma-1} H^{\gamma-1}.$$

Since all terms are positive except possibly $\rho - \gamma$, the sign of $\frac{\partial^2 Y}{\partial S \partial H}$ is determined by the sign of $\rho - \gamma$. Note that smaller ρ indicates greater diminishing returns, whereas smaller γ indicates greater complementarity between H and S in their aggregation.[24] Thus, S and H are *direct substitutes* in promoting achievement ($\frac{\partial^2 Y}{\partial S \partial H} < 0$) when diminishing returns dominate complementarity ($\rho < \gamma$), and *direct complements* otherwise.

We next consider the household's optimization problem. Given the exogenously determined level of school-tutoring time S, caregivers optimally choose home-tutoring time H to maximize their valuation of child achievement net of the cost of home-tutoring efforts.[25] For ease of illustration, we normalize the unit cost of home-tutoring time to be 1 for all households, but introduce heterogeneity by allowing the distribution of valuation of child achievement to vary by caregiver type τ (i.e., between grandparents and parents).[26] Therefore, the caregiver of child i's objective function in period $t \in \{0, 1\}$ (baseline and endline, respectively) is as follows:

$$\max_{H_{it}} \omega_{i\tau} E[Y_{ijt}] - \psi 1(H_{it} > 0) - H_{it}, \tag{9}$$

where $\omega_{i\tau} > 0$ stands for the caregiver's valuation of child achievement and is assumed to follow a Gamma distribution $\Gamma(\nu_\tau, \zeta_\tau)$, in which both the shape parameter ν_τ and the scale parameter ζ_τ differ by caregiver type τ. It is worth noting that when the shape parameter ν_τ is small, as is the case in our context, the Gamma distribution is a left-leaning curve with a mass distribution on the left tail. $\psi > 0$ is the fixed cost (in terms of time) for providing home tutoring, which, together with a small ν_τ, allows us to capture the fact that a large proportion

[24] In the extreme case in which $\gamma = -\infty$, S and H are perfect complements in producing T such that $T = min\{S^\rho, \theta H^\rho\}$.

[25] For simplicity, the model only considers one-child families. An extended model with more than one child will predict a stronger substitution effect within the household compared to the current model with one child. In multiple-children families, caregivers may spend more time with tutees' siblings without school tutoring after school tutoring becomes available to tutees.

[26] Note that here we normalize the unit cost of caregiver's time to be one and allow the distribution of the valuation of child achievement to vary by caregiver type. An alternative specification would be to normalize the valuation of children's achievement to be one and allow the unit cost to vary both between and within caregiver types. These two optimization problems are equivalent. The former can be transformed into the latter by dividing the objective function by the caregiver-specific valuation of child achievement $\omega_{i\tau}$: $Y_{ijt} - \frac{\psi}{\omega_{i\tau}} 1(H_{it} > 0) - \frac{1}{\omega_{i\tau}} H_{it}$. Therefore, for ease of illustration, we choose to normalize the unit time cost to be one and allow the distribution of the valuation of children's achievement to vary by caregiver type.

of caregivers choose zero home-tutoring time.[27] In a robustness check presented in Subsection 6.2, we allow ψ to vary by caregiver's type, but we find no evidence that fixed costs differ between parents and grandparents.

Given the additive linear specification of achievement production in Equation (7), we can omit the components of determining contemporary achievement $E[Y_{ijt}]$ that are irrelevant to home-tutoring time H_{it} (i.e., αY_{ijt-1}, δ_{jt}, and ϵ_{ijt}) and treat the caregiver's optimization problem as maximizing the return to total tutoring T_{it} net of the total cost of home-tutoring provision. The former depends on the product of the caregiver's valuation of child achievement $\omega_{i\tau}$ and the efficacy of total tutoring in achievement production A_τ, where the latter depends on the sum of the fixed and variable costs of home-tutoring time. Define $\kappa_{i\tau} = \omega_{i\tau} A_\tau$ as the caregiver-specific return to total tutoring T_{it}; the caregiver's objective function in Equation (9) can be simplified as

$$\max_{H_{it}} \kappa_{i\tau}\big[S_{it}^{\gamma} + \theta H_{it}^{\gamma}\big]^{\frac{\rho}{\gamma}} - \psi 1(H_{it} > 0) - H_{it}. \tag{10}$$

With the objective function Equation (10), the caregiver's interior solution for optimal home-tutoring time $H_{it}^{*}(\kappa_{i\tau}, S_{it}; \rho, \gamma, \theta)$ is given by

$$\rho\theta\big[S_{it}^{\gamma} + \theta H_{it}^{*^{\gamma}}\big]^{\frac{\rho}{\gamma}-1} H_{it}^{*^{\gamma-1}} = \frac{1}{\kappa_{i\tau}}. \tag{11}$$

As shown in Online Appendix D.1, applying the implicit-function theorem to Equation (11) yields the following relationship between the household's interior solution for optimal home-tutoring time H^* and school-tutoring time S:

$$\frac{dH^*}{dS} = \frac{(\rho-\gamma)S^{\gamma-1}}{(1-\rho)\theta H^{*^{\gamma-1}} + (1-\gamma)S^{\gamma} H^{*^{-1}}}. \tag{12}$$

Since all terms in Equation (12) are positive except possibly $\rho - \gamma$, $\frac{dH^*}{dS}$ has the same sign as $\rho - \gamma$. In particular, when $\rho < \gamma$, that is, S and H are *direct substitutes* in achievement production, $\frac{dH^*}{dS} < 0$.

Although the interior solution for optimal home-tutoring time H_{it}^* given by Equation (11) is always positive, the caregiver would still choose zero home-tutoring time if it yields a higher net value than choosing H_{it}^*, that is, $\kappa_{i\tau} S_{it}^{\rho} > \kappa_{i\tau}[S_{it}^{\gamma} + \theta H_{it}^{*^{\gamma}}]^{\frac{\rho}{\gamma}} - \psi - H_{it}^*$, which is possible with a fixed cost of home tutoring ψ. Therefore, the unconstrained optimal home-tutoring time, denoted as H_{it}^{**}, is given by

$$H_{it}^{**} = \begin{cases} H_{it}^{*} & \text{if} \quad \kappa_{i\tau}\big[S_{it}^{\gamma} + \theta H_{it}^{*^{\gamma}}\big]^{\frac{\rho}{\gamma}} - \psi - H_{it}^{*} > \kappa_{i\tau} S_{it}^{\rho} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

In the remainder of this section, we derive some propositions generated from the model under the condition $\rho < \gamma$ (i.e., S and H are *direct substitutes* in achievement production), which will be examined in the structural estimation in the next section. We provide some intuitions for both interior and corner solutions in the main text. More detailed discussions about the interior and corner solutions are in Online Appendix D and E, respectively.

PROPOSITION 1. *Households' optimal home-tutoring times decrease weakly with school-tutoring times.*

[27] The idea of introducing fixed costs to allow for corner solutions is common in the trade literature. For example, Helpman et al. (2008) introduce fixed costs of exporting such that some firms choose not to export.

When we consider the interior solution of home-tutoring choice, under the condition $\rho < \gamma$, households' optimal home-tutoring times decrease with school-tutoring times, which is derived directly from Equation (12). When allowing for corner solutions, a weaker version holds in the sense that households' optimal home-tutoring times decrease *weakly* with school-tutoring times. In particular, there exists a household-specific threshold value $S^* \geq 0$ such that the caregiver chooses zero home-tutoring time when $S \geq S^*$. The detailed proofs of interior and corner solutions are in Online Appendix D.1 and E.1, respectively.

Given Proposition 1, under the interior solution, although the direct effect of increased school-tutoring time on total tutoring T ($\frac{\partial T}{\partial S}$) is positive, the indirect effect through household's substitution in home-tutoring time ($\frac{\partial T}{\partial H^*} \cdot \frac{dH^*}{dS}$) is negative. Nonetheless, as we show formally in Online Appendix D.2, when caregivers choose the interior solution (positive home-tutoring time), the positive direct effect always dominates the negative indirect effect such that the net effect on total tutoring ($\frac{dT}{dS} = \frac{\partial T}{\partial S} + \frac{\partial T}{\partial H^*} \cdot \frac{dH^*}{dS}$) remains positive despite household's substitution in home-tutoring time. Given the achievement-production function, Equation (7), a net increase in total tutoring T also leads to an increase in achievement Y. When we allow for the corner solution, as shown in Online Appendix E.2, increasing school-tutoring time has a net positive effect on both total tutoring and student achievement almost everywhere except for the point where the increase in school-tutoring time happens to induce households to switch from choosing positive home-tutoring time to zero home-tutoring time.

Proposition 2. *The positive direct effect of increased school-tutoring time dominates the negative indirect effect through household's optimal substitution in home-tutoring time, resulting in an overall increase in both total tutoring and student achievement almost everywhere except for the point where the increase in school-tutoring time happens to induce households to switch from choosing positive home-tutoring time to zero home-tutoring time.*

Equation (11) shows that in addition to school-tutoring time S_{it}, the optimal choice of home-tutoring time H^*_{it} also depends on the caregiver-specific returns to total tutoring $\kappa_{i\tau}$. As shown in Online Appendix D.3, applying the implicit-function theorem to Equation (11) shows that given school-tutoring time S, under the parameter space considered here (i.e., $\kappa > 0, \theta > 0$, and $0 < \rho < \gamma < 1$), household's optimal interior solution for home-tutoring time H^* increases with caregiver's returns to total tutoring κ, that is,

$$\frac{dH^*}{d\kappa} = \frac{1}{\kappa^2 \rho\theta[S^\gamma + \theta(H^*)^\gamma]^{\frac{\rho}{\gamma}-2}(H^*)^{\gamma-2}((1-\gamma)S^\gamma + (1-\rho)\theta(H^*)^\gamma)} > 0.$$

When allowing for corner solutions, a weaker version holds such that home-tutoring time increases weakly with caregiver-specific returns to total tutoring κ conditional on fixed costs of home-tutoring ψ and school-tutoring time S. In particular, for a given fixed cost of home-tutoring provision ψ and a given level of school-tutoring time S, there exists a threshold value $\kappa^*(S, \psi)$ such that caregivers with $\kappa > \kappa^*(S, \psi)$ choose positive home-tutoring time whereas those with $\kappa \leq \kappa^*(S, \psi)$ choose zero home-tutoring time. More detailed discussions of corner solutions are provided in Online Appendix E.3.

Proposition 3. *Given the same level of school-tutoring time, children cared for by caregivers with higher κ receive a weakly higher level of home-tutoring time than those cared for by caregivers with lower κ.*

We next consider how the adjustments in the optimal level of home-tutoring time (i.e., $\frac{dH^*}{dS}$) differ across caregivers with different returns to total tutoring κ. In Online Appendix D.4, we derive formally that the cross-derivative $\frac{d^2H^*}{dSd\kappa}$ is negative at interior solutions. This together with $\frac{dH^*}{dS} < 0$ (Proposition 1), implies that when school-tutoring time increases, caregivers with higher κ reduce their home-tutoring time more than those with lower κ. When accounting for corner solutions, as shown in Online Appendix E.4, a weaker version holds except for the point where the caregiver with a lower κ switches from the interior solution to the corner solution.

PROPOSITION 4. *When school-tutoring time increases, caregivers with higher κ reduce their home-tutoring time weakly more than those with lower κ almost everywhere except for the point where the caregiver with a lower κ switches from the interior solution to the corner solution.*

We consider how the net effect of increased school-tutoring time on total tutoring (i.e., $\frac{dT}{dS}$) differs across children cared for by caregivers with different κ. We derive formally in Online Appendix D.5 that the cross-derivative $\frac{d^2T}{dSd\kappa}$ is negative at interior solutions, which together with $\frac{dT}{dS} > 0$ (Proposition 2) implies that an increase in school-tutoring time leads to a smaller increase in total tutoring for children cared for by caregivers with higher κ than those cared for by caregivers with lower κ.[28] When considering corner solutions, a weaker version holds almost everywhere except for the point where the household with lower κ switches from the interior solution to the corner solution. Online Appendix E.5 provides more discussions regarding corner solutions.

PROPOSITION 5. *When school-tutoring time increases, children cared for by caregivers with higher κ experience a weakly smaller increase in total tutoring T than those cared for by caregivers with lower κ almost everywhere except for the point where the caregiver with a lower κ switches from an interior solution to a corner solution.*

Figure 3 summarizes the predictions of our theoretical analysis (under the condition $\rho < \gamma$) regarding the optimal interior choice of home-tutoring time for two representative households with children cared for by parents and grandparents, denoted by p and g, respectively. In line with our empirical observation that parents on average choose higher levels of home-tutoring time than grandparents, we assume that the representative parents have higher returns to total tutoring than the representative grandparents, $\bar{\kappa}_p > \bar{\kappa}_g$. In the figure, the vertical axis represents the left-side expression of Equation (11), that is, $\frac{dT}{dH}$, and the two horizontal lines correspond to the, respective, right-side values of Equation (11) for the representative parents and grandparents, that is, the inverse of their valuation of total tutoring $\frac{1}{\bar{\kappa}_p}$ and $\frac{1}{\bar{\kappa}_g}$. First, with an initial level of school-tutoring time S, parents choose a higher optimal level of home-tutoring time than grandparents, $\bar{H}_p^* > \bar{H}_g^*$ (Proposition 3). Second, when the level of school-tutoring time increases from S to S', both parents and grandparents reduce their level of home-tutoring time, $\bar{H}_p^{*\prime} < \bar{H}_p^*$ and $\bar{H}_g^{*\prime} < \bar{H}_g^*$ (Proposition 1). Third, the extent of the reduction in home-tutoring time is greater for parents than grandparents, $\bar{H}_p^* - \bar{H}_p^{*'} > \bar{H}_g^* - \bar{H}_g^{*'}$ (Proposition 4).

Although not demonstrated in Figure 3, Proposition 2 implies that the positive direct effect of increased school-tutoring time dominates the negative indirect effect of decreased home-tutoring time such that higher school-tutoring time leads to higher total tutoring for both types of caregivers, that is, $\bar{T}_p' > \bar{T}_p$ and $\bar{T}_g' > \bar{T}_g$. Moreover, given that $\bar{\kappa}_g < \bar{\kappa}_p$, Proposition 5 further predicts a larger increase in total tutoring for children cared for by grandparents than those cared for by parents, that is, $\Delta\bar{T}_g = \bar{T}_g' - \bar{T}_g > \Delta\bar{T}_p = \bar{T}_p' - \bar{T}_p$. However, whether and to what extent the larger increase in T for children cared for by grandparents translates into larger achievement gains also depends on the relative magnitude of the two efficacy parameters A_g and A_p. Given that $\Delta\bar{Y}_\tau = A_\tau \Delta\bar{T}_\tau$ for $\tau \in \{p, g\}$, a sufficient condition for children

[28] Note that Proposition 5 only states a monotone relationship between changes in total tutoring ΔT and caregiver-specific returns to total tutoring κ but not one between achievement gains ΔY and κ. As discussed later in this section, the former does not necessarily translate into the latter because children may also differ in the efficacy of total tutoring in achievement production A.

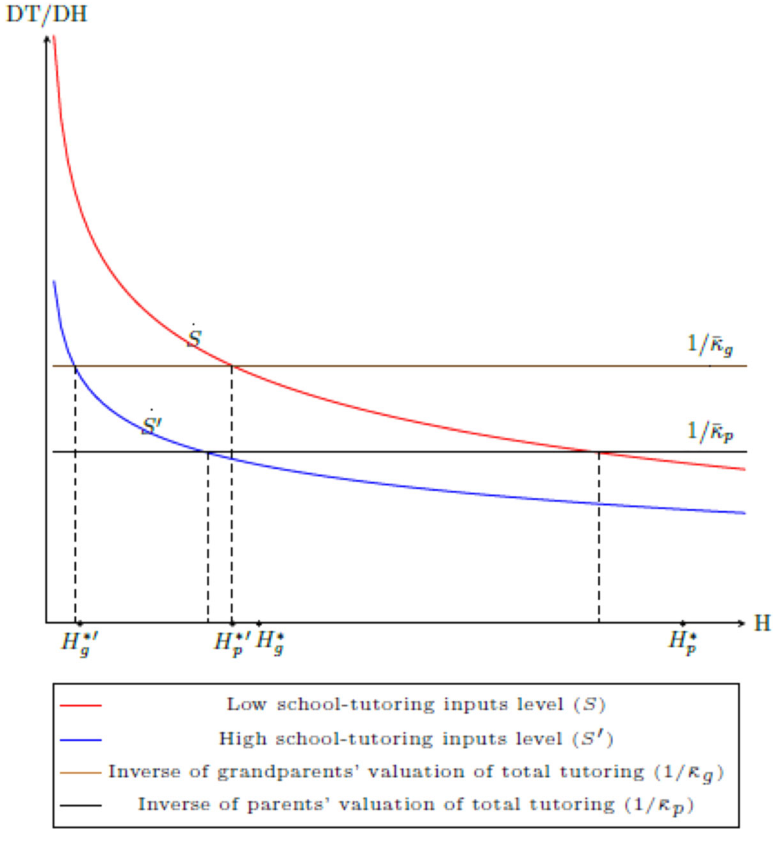


FIGURE 3

HOUSEHOLD'S OPTIMAL CHOICES OF HOME-TUTORING TIME

cared for by grandparents to experience larger achievement gains than those cared for by parents is $A_g > A_p$, under which differences in both A_τ and $\Delta\bar{T}_\tau$ will lead to $\Delta\bar{Y}_g > \Delta\bar{Y}_p$. Denote $\bar{\omega}_p$ and $\bar{\omega}_g$ the, respective, valuation of child achievement of the representative parents and grandparents. Given that $\bar{\kappa}_\tau = A_\tau\bar{\omega}_\tau$ for $\tau \in \{p, g\}$, under the aforementioned assumption $\bar{\kappa}_p > \bar{\kappa}_g$, the condition $A_g > A_p$ holds if the difference in the valuation of child achievement between parents and grandparents dominates the difference in the efficacy parameter, that is, $\frac{\bar{\omega}_p}{\bar{\omega}_g} > \frac{A_g}{A_p} > 1$.

6. ESTIMATION

In this section, we first describe our estimation procedure (Subsection 6.1), then report our estimates (Subsection 6.2), and finally present decomposition and counterfactual analysis (Subsection 6.3).

6.1. *Estimation Procedure.* We employ simulated maximum likelihood (SML) to jointly estimate the parameters in the achievement-production function ($A_\tau, \rho, \gamma, \theta, \alpha$, and σ) and those in the caregiver's maximization problem (ν_τ, ζ_τ, and ψ) using data of tutees and controls in the baseline and endline. Because controls report a larger decline in home-tutoring time compared to tutees in the endline survey, we allow for potential underreporting of home tutoring for controls who actually received tutoring at home in the endline. Suppose H_{i1}^{**} is the *actual* home-tutoring time for control i in the endline. We assume when optimal home-tutoring time is positive, with probability p_u, the control respondent underreported his/her home-tutoring time by l_u. Hence, the reported home-tutoring time is linked to the *actual* home-tutoring time as follows:

$$(14)\quad H_{it} = \begin{cases} \mathbf{1}\{r_i \geq p_u\}H_{it}^{**} + \mathbf{1}\{r_i < p_u\}\max\{H_{it}^{**} - l_u, 0\} & i \in \text{controls}, t = 1, H_{it}^{**} > 0, \\ H_{it}^{**} & \text{otherwise}, \end{cases}$$

where r_i is a random number drawn from the uniform distribution. For control respondents with positive optimal home-tutoring time in the endline, if $r_i \geq p_u$, caregivers report the *actual* home-tutoring time; otherwise, they underreport home-tutoring time by l_u. If $H_{it}^{**} - l_u < 0$, we assume that the caregivers report zero home-tutoring time.

We estimate p_u and l_u together with other parameters in the model to detect the underreporting problem. If $p_u = 0$, there is no underreporting behavior for controls. For tutees and controls in the baseline, we assume no underreporting problem. In addition, we assume no trend in home-tutoring input. Therefore, the underreporting parameters are identified from the lack of fit for endline home inputs among the control group between the model and data. In particular, p_u is identified by the difference in the share of children with zero home tutoring between the data and model for controls in the endline, and l_u is identified by the difference in the average home-tutoring time between the data and model for the same group of people.

Since we observe school- and home-tutoring time, as well as baseline and endline test scores for tutees, the parameters in the production function can be identified. The distribution of home-tutoring time, including the share of children with zero home tutoring and the average home-tutoring time, identifies caregivers' valuation of child achievement and fixed costs of home tutoring.

The likelihood contains two parts. The first part is the likelihood of endline achievement and the second is the likelihood of baseline and endline home-tutoring time. The log likelihood is

$$L = \sum_i \log(\phi_y(Y_{ij1})) + \sum_i \sum_{t \in \{0,1\}} \log(\phi_h(H_{it})),$$

where ϕ_y is the probability density of endline score Y_{ij1} and ϕ_h is the probability density of *observed* baseline and endline home-tutoring time H_{it}. We match the likelihood for both tutees and controls.

From Equations (7) and (8), the observed endline-math score is given by

$$Y_{ij1} = \alpha Y_{ij0} + A_\tau [S_{i1}^\gamma + \theta (H_{i1}^{**})^\gamma]^{\frac{\rho}{\gamma}} + \delta_{j1} + \epsilon_{ij1},$$

The school-tutoring time S_{it} is set to be three hours per week for tutees and zero for controls. H_{i1}^{**} is the *actual* home-tutoring time in the endline. For tutees, we use their *self-reported* home-tutoring time in the endline as the *actual* home-tutoring time. For controls, since they probably underreport their home-tutoring time in the endline, we simulate their *actual* home-tutoring time from their caregivers' optimization problem. In particular, we draw caregivers' valuations of child achievement $\omega_{i\tau}$ from the Gamma distribution $\Gamma(\nu_\tau, \zeta_\tau)$ and predict their *actual* home-tutoring time numerically using Equation (13). We repeat the simulation 100 times for each control to get their average *actual* home-tutoring time.

We assume that ϵ_{ij1} is an i.i.d. random variable drawn from a normal distribution $N(0, \sigma^2)$ orthogonal to home-tutoring time H_{i1}. The class fixed effects δ_{j1} are estimated by taking the class-level average on both sides of Equation (7), that is, $\delta_{j1} = E_j[Y_{ij1}] - \alpha E_j[Y_{ij0}] + A_\tau E_j[(S_{i1}^\gamma + \theta(H_{i1}^{**})^\gamma)^{\frac{\rho}{\gamma}}]$, where $E_j[\cdot]$ represents taking the class-level average. The probabil-

TABLE 7
ESTIMATION RESULTS OF THE STRUCTURAL MODEL

Panel A. Production Function

	ρ	γ	θ	A_g	A_p	α	σ
Coeff.	0.147	0.223	0.673	0.288	0.163	0.571	0.693
S.E.	(0.015)	(0.037)	(0.044)	(0.029)	(0.017)	(0.278)	(0.290)
	$\gamma - \rho$	$A_g - A_p$	$\theta - 1$				
Diff.	0.076**	0.126***	−0.327***				
S.E.	(0.038)	(0.033)	(0.044)				

Panel B. Optimization Problem

	ν_g	ν_p	ζ_g	ζ_p	ψ	p_u	l_u
Coeff.	0.080	0.110	448.9	929.7	0.505	0.518	1.664
S.E.	(0.008)	(0.021)	(45.8)	(109.4)	(0.085)	(0.077)	(0.329)

NOTE: We employ the sample of both tutees and controls in the baseline and endline periods and use the maximum likelihood estimation (MLE) to estimate the parameters of the achievement production function (Panel A) and household optimization problem (Panel B). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

ity density of Y_{ij1} is obtained by $\phi_n((Y_{ij1} - \hat{Y}_{ij1})/\sigma)/\sigma$, where ϕ_n is the probability density of a standard normal distribution and $\hat{Y}_{ij1}$ is the model's predicted test score.

For the second part of the likelihood, we numerically simulate the distribution of *observed* home-tutoring time for tutees and controls in both periods $t \in \{0, 1\}$. School-tutoring time is set to zero hours for controls, zero hours for tutees in the baseline, and three hours for tutees in the endline. We simulate the optimal home-tutoring time by drawing caregivers' valuations of child achievement $\omega_{i\tau}$ from the Gamma distribution $\Gamma(\nu_\tau, \zeta_\tau)$ and predict their optimal home-tutoring time numerically following Equation (13). For tutees, controls in the baseline, and controls in the endline with zero optimal home tutoring, we assume that their *observed* home-tutoring time is equal to their *actual* home-tutoring time ($H_{i1} = H_{i1}^{**}$). For controls with positive optimal home-tutoring time in the endline, we simulate their observed home-tutoring time by accounting for the underreporting behavior following Equation (14). In particular, for each individual, we make a draw of r_i from the uniform distribution. If $r_i < p_u$, the individual reports zero home-tutoring time if $H_{it}^{**} \leq l_u$ and reports $H_{it}^{**} - l_u$ if $H_{it}^{**} > l_u$, that is, $H_{it} = \max\{H_{it}^{**} - l_u, 0\}$. If $r_i \geq p_u$, the individual reports the *actual* home-tutoring time ($H_{i1} = H_{i1}^{**}$). We repeat the simulation 500 times for the whole sample to obtain the numerical distribution of H_{it}, which allows for the corner solution.

6.2. *Estimates.* Panel A of Table 7 shows the estimates for the achievement production function, Equation (7). The estimated ρ and γ are 0.147 and 0.223, respectively. Their difference is statistically significant at the 5% level. This confirms $\rho < \gamma$, the condition that we impose in deriving the model predictions in Section 5, implying that diminishing returns dominate complementarity and school tutoring and home tutoring are direct substitutes. The estimated efficiency for home tutoring θ is 0.673, suggesting that home-tutoring time is two-thirds as efficient as school-tutoring time, whose factor loading is normalized to be one. The estimated A_g (0.288) is larger than A_p (0.163) and their difference is statistically significant at the 1% level, suggesting that tutoring is more efficacious for children growing up in disadvantaged early life circumstances (i.e., left-behind by parents in care of grandparents). Our finding of the substitutability between early life circumstances and contemporaneous tutoring inputs is consistent with several recent studies documenting that latter life investments can help remediate disadvantages generated very early in life (Adhvaryu et al., forthcoming; Goff et al., forthcoming; Rossin-Slater and Wüst, 2020).[29]

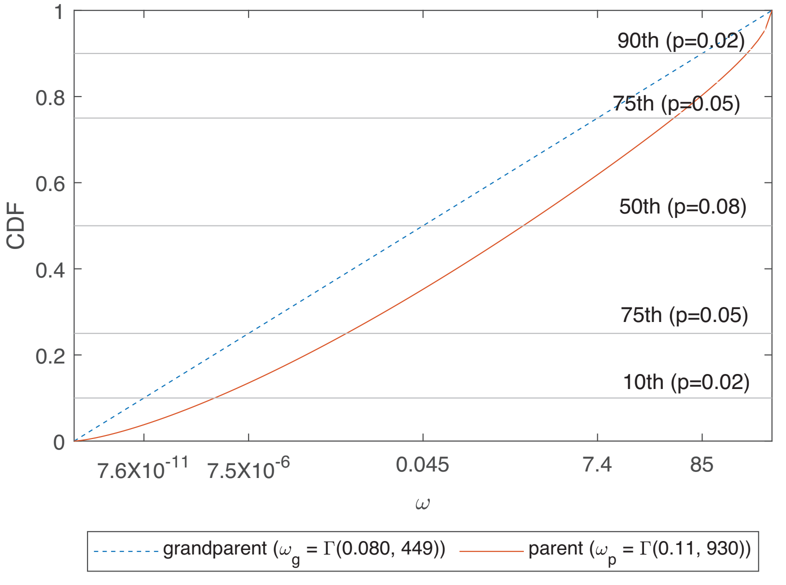
Panel B of Table 7 further presents the estimates for households' optimization problem, Equation (10). The valuation of child achievement $\omega_{i\tau}$ follows the Gamma distribution with the shape parameter ν_τ and scale parameter ζ_τ for $\tau = \{g, p\}$. Panel A of Figure 4 plots the respective cumulative distribution function (CDF) curves of ω for grandparents and parents. Since the estimated shape parameters are small for both parents and grandparents (0.080 and 0.110, respectively), both distributions are highly positively skewed with a mass distribution on the left tail. For ease of illustration, we use the percentile scale of ω for grandparents in the horizontal axis of Panel A such that the CDF curve for grandparents is the 45-degree line from the origin. The CDF curve of ω for parents is always to the right of the 45-degree line, showing that $F_p^{-1}(q) > F_g^{-1}(q)\ \forall q$, where $F_\tau^{-1}(.)$ denotes the inverse CDF function of ω for $\tau \in \{p, g\}$ and q denotes percentile. Next, we further check whether the quantitative differences in the CDF curves shown in Panel A of Figure 4 provide statistical support for the distribution of ω for parents to first-order stochastically dominate (FOSD) that for grandparents. To do so, we employ a bootstrap strategy that calculates the one-sided p-value for the observed difference in ω between parents and grandparents at percentile q under the null hypothesis $F_p^{-1}(q) \leq F_g^{-1}(q)$ for $q \in \{10, 25, 50, 75, 90\}$. The implementation details of the bootstrap process are documented in the Online Appendix C.2. The p-values for the bootstrap test are 0.02, 0.05, 0.08, 0.05, and 0.02 for the 10th, 25th, 50th, 75th, and 90th percentiles, respectively, suggesting that the differences in the CDF curves in Panel A of Figure 4 at these percentiles are statically significant at the 10% level.

Although parents FOSD grandparents in terms of valuation of child achievement ($\omega_{i\tau}$), it remains unclear who has higher returns to total tutoring ($\kappa_{i\tau} = A_\tau \omega_{i\tau}$) because $A_p < A_g$. In order to answer this question, we scale the two CDF curves ω_p and ω_g in Panel A of Figure 4 horizontally by the corresponding estimates of the efficacy parameter of total tutoring (i.e., 0.163 for ω_p and 0.288 for ω_g) to obtain the CDF curves for κ_p and κ_g in Panel B of Figure 4. The κ_p curve is always to the right of the κ_g curve, suggesting that parents have higher κ than grandparents at any percentile of the distribution. Moreover, the bootstrap test yields p-values of 0.04, 0.07, 0.09, 0.06, and 0.04 for the 10th, 25th, 50th, 75th, and 90th percentile, respectively, suggesting that the differences in κ between parents and grandparents at these percentiles are statistically significant at the 10% level. Given Propositions 3 and 4 in Section 5, when $\kappa_p > \kappa_g$, non-LBC would receive higher levels of home-tutoring time than LBC and increases in school-tutoring time leads to larger reductions in home-tutoring time for non-LBC than LBC, both of which are confirmed by the empirical results in Table 6.

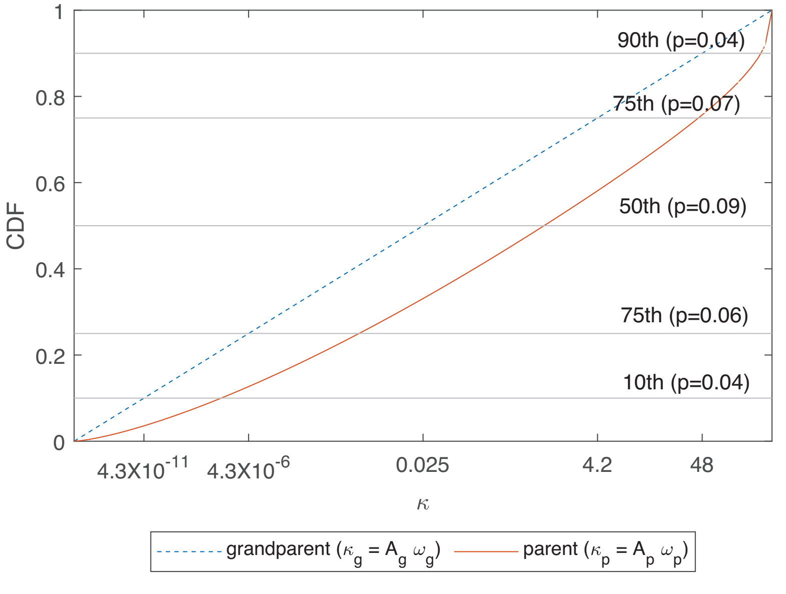
Finally, the estimation shows that the fixed cost $\psi = 0.505$. This, together with the mass distribution of ω on the left tail, allows the model to predict a large share of households choosing zero home-tutoring time as observed in the data.

Panel A of Table 8 presents the model fit for endline test scores of controls and tutees by caregiver types. Similar to Subsection 4.4, we adjust for differences in treatment probabilities across classes using inverse probability weights that equal to $1/\bar{D}_j$ for tutees and $1/(1 - \bar{D}_j)$ for controls, where $\bar{D}_j$ is the probability of treatment of class j. First, the model matches the endline test scores for LBC controls (column 1), LBC tutees (column 2), non-LBC controls (column 4), and non-LBC tutees (column 5). Second, the model predicts substantially higher endline scores for tutees than controls, and the predicted score differences between tutees

[29] Specifically, Adhvaryu et al. (forthcoming) find that the PROGRESA program in Mexico had the largest effects on children impacted by negative rainfall shocks at birth with each additional year of program exposure mitigating almost 20% of early disadvantage; Rossin-Slater and Wüst (2020) interact preschool access at age 3 with access to nurse home visit program in infancy in Denmark and find that access to nurse home visit program reduced the positive impact of preschool on the human capital index by 74%; Goff et al. (forthcoming) also find some suggestive evidence that access to better high school yields smaller gains in Baccalaureate exam scores among Romanian students who were born after the lifting of an abortion ban and thus experienced better family environments and greater parental investments during childhood.



(a) Panel A: CDF of ω



(b) Panel B: CDF of κ

NOTES: The x-axis for Panel A (Panel B) is the percentile of ω (κ) for grandparents, so the CDF of ω (κ) for grandparents is on the 45-degree line.

FIGURE 4

CDF OF ω AND κ FOR GRANDPARENTS AND PARENTS

and controls are significantly larger for LBC tutees (0.159σ, column 2) than non-LBC tutees (0.073σ, column 5).

Panels B and C of Table 8 report the model fit on the share choosing positive home-tutoring time and average weekly home-tutoring hours for tutees, respectively. In the baseline period, the model predicts that non-LBC tutees are more likely to receive positive home tutoring and

Table 8
MODEL FIT

	LBC			Non-LBC		
	Control (1)	Tutee (2)	Tutee - Control (3)	Control (4)	Tutee (5)	Tutee - Control (6)
Panel A. Endline Test Scores						
Data	−0.592	−0.426	0.166	−0.465	−0.393	0.072
Model	−0.592	−0.433	0.159	−0.464	−0.392	0.073

	LBC			non-LBC		
	Baseline (1)	Endline (2)	Endline - Baseline (3)	Baseline (4)	Endline (5)	Endline - Baseline (6)
Panel B. Share Choosing Positive Home-Tutoring Time, Tutees						
Data	0.230	0.133	−0.097	0.375	0.251	−0.124
Model	0.256	0.186	−0.070	0.346	0.248	−0.098
Panel C. Average Weekly Home-Tutoring Hours, Tutees						
Data	0.802	0.507	−0.295	1.557	1.123	−0.434
Model	0.804	0.594	−0.210	1.422	1.084	−0.338
Panel D. Share Choosing Positive Home-Tutoring Time, Controls						
Data	0.285	0.135	−0.150	0.350	0.189	−0.161
Model	0.256	0.138	−0.118	0.346	0.213	−0.133
Panel E. Average Weekly Home-Tutoring Hours, Controls						
Data	1.070	0.558	−0.511	1.550	0.897	−0.653
Model	0.804	0.480	−0.324	1.422	1.041	−0.380

Note: Panel A compares the observed data and model predictions for endline test scores of controls and tutees by caregiver type. Panels B and C (D and E) compare the observed data and model predictions for the share choosing positive home-tutoring time and average weekly home-tutoring hours for LBC and non-LBC tutees (controls). All panels adjust for differences in treatment probabilities across classes using inverse probability weights.

have higher average home-tutoring hours than LBC tutees. With the provision of school tutoring in the endline period, the model predicts the home-tutoring time to decline for both types of tutees but to a greater extent for non-LBC tutees. All these predictions are in line with observed data reported in the upper rows and the reduced-form results on home-tutoring inputs in Table 6.

The model fits on the share choosing positive home-tutoring time and average weekly home-tutoring hours for controls are further reported in Panels D and E of Table 8, respectively. The model's predicted baseline home-tutoring for controls is the same as that for tutees because there is no systematic difference in the model that will lead to differential responses between the two groups. The model's predicted endline home tutoring for controls has been adjusted for underreporting and it matches the observed home tutoring (in the upper rows) well. Table 7 shows that 51.8% of controls with positive optimal home-tutoring time underreport and they underreport on average by 1.66 hours per week. Since only 30.6% of controls have positive optimal home-tutoring time, this suggests that 15.8% (30.6% $\times$ 51.8%) of controls underreport their home-tutoring time.

In Online Appendix F, we conduct a number of robustness checks on the estimations of the achievement production function and household's optimization problem. First, we check whether the efficiency of home-tutoring time (θ) varies by caregiver's type or education. The estimated differences in θ are never statistically significant, suggesting no statistical evidence for heterogeneity in home-tutoring effectiveness by caregiver's type or education. Second, we check whether the efficacy of total tutoring (A) varies by school quality as proxied by the class-level value added in test scores among control students, and find no evidence that school quality has any significant effect on A. Third, we instrument baseline math scores with baseline reading scores to deal with the measurement errors in baseline math scores. We find similar estimation results in this robustness check. Fourth, in addition to the valuation of child achievement, we also allow the fixed cost ψ to vary by caregiver type. The estimation results show that grandparents have higher fixed costs than parents, but the difference is not statistically significant. Finally, we further allow the parameters in the valuation of child achievement (ν and ζ) to vary by caregiver's education conditional on caregiver's type and find no evidence for heterogeneity in the valuation of child achievement by caregiver's education conditional on caregiver's type.

6.3. *Decomposition and Counterfactual Analysis.* We first decompose the difference in test scores between LBC and non-LBC children into the difference in their home-tutoring time and the difference in the related tutoring efficiencies. We then undertake counterfactual analysis to examine how optimal choices of home-tutoring time and marginal treatment effects (MTEs) of school tutoring vary by caregiver's type and valuation of student achievement and next compare the current school-tutoring program with equal school-tutoring time for LBC and non-LBC students with an alternative program that optimally allocates tutoring time between LBC and non-LBC students.

For each student i of type τ ($\in \{g, p\}$), we can use the estimated parameters to simulate the expected optimal home-tutoring time and test scores with the school-tutoring intervention ($H^{**}_{i\tau 1}$ and $\hat{Y}_{i\tau 1}$) by setting $S = 3$, and the counterfactual home-tutoring time and test scores ($H^{**}_{i\tau 0}$ and $\hat{Y}_{i\tau 0}$) by setting $S = 0$. The individual-specific treatment effect predicted by the model, denoted as $\Phi_{i\tau}$, can be expressed as the difference between the two simulated test-score changes:

$$\Phi_{i\tau} = \hat{Y}_{i\tau 1} - \hat{Y}_{i\tau 0} = A_\tau[3^\gamma + \theta(H^{**}_{i\tau 1})^\gamma]^{\rho/\gamma} - A_\tau[\theta(H^{**}_{i\tau 0})^\gamma]^{\rho/\gamma}.$$

Aggregating the simulated individual-specific treatment effects across all students of type τ yields the predicted CATE for students of type τ, denoted as $E_{A_\tau}(\Phi_{i\tau}|\boldsymbol{H^{**}_{i\tau}})$, as follows:

$$E_{A_\tau}(\Phi_{i\tau}|\boldsymbol{H^{**}_{i\tau}}) = \frac{\sum_{i=1}^{N_\tau}\Phi_{i\tau}}{N_\tau},$$

where $\boldsymbol{H^{**}_{i\tau}} = \{H^{**}_{i\tau 0}, H^{**}_{i\tau 1}\}$, consisting of student i's home-tutoring time both before and after the treatment, and N_τ is the number of students of type τ.

Following Oaxaca and Ransom (1994), we employ a generalized Oaxaca–Blinder decomposition of the difference in the predicted CATE between LBC tutees and non-LBC tutees as follows:

$$\begin{aligned} & E_{A_g}(\Phi_i|\boldsymbol{H^{**}_{ig}}) - E_{A_p}(\Phi_i|\boldsymbol{H^{**}_{ip}}) \\ (15) \qquad & = \Big(E_{\bar{A}}(\Phi_i|\boldsymbol{H^{**}_{ig}}) - E_{\bar{A}}(\Phi_i|\boldsymbol{H^{**}_{ip}})\Big) \\ & + \Big([E_{A_g}(\Phi_i|\boldsymbol{H^{**}_{ig}}) - E_{\bar{A}}(\Phi_i|\boldsymbol{H^{**}_{ig}})] + [E_{\bar{A}}(\Phi_i|\boldsymbol{H^{**}_{ip}}) - E_{A_p}(\Phi_i|\boldsymbol{H^{**}_{ip}})]\Big), \end{aligned}$$

where $\bar{A} = 0.5A_g + 0.5A_p$, the average tutoring efficacy parameter between LBC and non-LBC, and $E_{\bar{A}}(\Phi_i|\boldsymbol{H^{**}_{ip}})$ ($E_{\bar{A}}(\Phi_i|\boldsymbol{H^{**}_{ig}})$) denotes the counterfactual CATE when the home-tutoring time (both before and after the treatment) of non-LBC (LBC) is evaluated at the

TABLE 9
DECOMPOSITION OF THE DIFFERENCE IN THE TREATMENT EFFECT BETWEEN LBC AND NON-LBC

Panel A. Simulated Treatment Effect		
LBC $E_{A_g}(\Phi_i\|\boldsymbol{X_{ig}})$	Non-LBC $E_{A_p}(\Phi_i\|\boldsymbol{X_{ip}})$	Diff. $E_{A_g}(\Phi_i\|\boldsymbol{X_{ig}}) - E_{A_p}(\Phi_i\|\boldsymbol{X_{ip}})$
0.238	0.123	0.115 (100%)
Panel B. Counterfactual CATE by Using $\bar{A}$		
LBC $E_{\bar{A}}(\Phi_i\|\boldsymbol{X_{ig}})$	Non-LBC $E_{\bar{A}}(\Phi_i\|\boldsymbol{X_{ip}})$	
0.186	0.171	
Panel C. Decomposition Analysis		
Diff. in H $E_{\bar{A}}(\Phi_i\|\boldsymbol{X_{ig}}) - E_{\bar{A}}(\Phi_i\|\boldsymbol{X_{ip}})$	Diff. in A_g and $\bar{A}$ $E_{A_g}(\Phi_i\|\boldsymbol{X_{ig}}) - E_{\bar{A}}(\Phi_i\|\boldsymbol{X_{ig}})$	Diff. in $\bar{A}$ and A_p $E_{\bar{A}}(\Phi_i\|\boldsymbol{X_{ip}}) - E_{A_p}(\Phi_i\|\boldsymbol{X_{ip}})$
0.015 (13.1%)	0.052 (45.3%)	0.048 (41.6%)

NOTE: Panel A reports the conditional average treatment effect (CATE) for LBC, non-LBC, and their difference predicted by the estimated model parameters. Panel B presents the simulated counterfactual CATE for LBC and non-LBC evaluated at the average efficacy parameter $\bar{A}$ for LBC. Panel C decomposes the difference in the CATE between LBC and non-LBC into that attributable to the difference in home-tutoring time and the difference in their tutoring efficacy.

average tutoring efficacy parameter. In Equation (15), the first term corresponds to the difference in the CATE between LBC and non-LBC attributable to the difference in their home-tutoring time ($\boldsymbol{H}_{ig}^{**}$ vs. $\boldsymbol{H}_{ip}^{**}$) evaluated at the average efficacy parameter ($\bar{A}$), the second term corresponds to that attributable to the difference between LBC's efficiency parameter and the average efficacy parameter (A_g vs. $\bar{A}$) evaluated at the home-tutoring time of LBC ($\boldsymbol{H}_{ig}^{**}$), and the last term corresponds to the difference associated with the difference between the average efficacy parameter and non-LBC's efficiency parameter ($\bar{A}$ vs. A_p) evaluated at the home-tutoring time of non-LBC ($\boldsymbol{H}_{ip}^{**}$).

Table 9 reports the decomposition results. Panel A shows that the model predicts a CATE of 0.238σ for LBC and 0.123σ for non-LBC, both of which are consistent with the estimates reported in Table 3. Panel B simulates a counterfactual CATE of 0.186σ (0.171σ) for LBC (non-LBC) by applying the average tutoring = efficacy parameter $\bar{A}$. Using these counterfactual CATEs as intermediates, Panel C decomposes the total difference in the CATE between LBC and non-LBC (0.115σ) into that attributable to their difference in changes in home-tutoring time (0.015σ or 13.1%), that attributable to the difference in LBC's tutoring efficiency and the average tutoring efficiency (0.052σ or 45.3%), and that attributable to the difference in the average tutoring efficiency and non-LBC's tutoring efficiency (0.048σ or 41.6%). The results of this decomposition analysis suggest that the difference in tutoring efficacy (A_g vs. A_p) accounts for 7/8s of the total difference in the CATE of the after-school tutoring intervention between LBC and non-LBC, whereas the difference in changes in home-tutoring time ($\boldsymbol{H}_{ig}^{**}$ vs. $\boldsymbol{H}_{ip}^{**}$) accounts for the remaining 1/8.

Next, we conduct a simulation exercise to examine how the optimal choice of home-tutoring time and MTE of school tutoring vary by caregiver's type and valuation of student achievement. Figure 5 compares the two types of caregivers with ω equal to the 55th, 65th, 75th, and 85th percentile of their respective distribution. As demonstrated in detail below, these percentiles are selected so that the respective panels of Figure 5 demonstrate four distinct scenarios for how parents' and grandparents' optimal choices of home-tutoring time (H^{**}) respond to increases in school-tutoring time (S). Panel A shows that, at the 55th per-

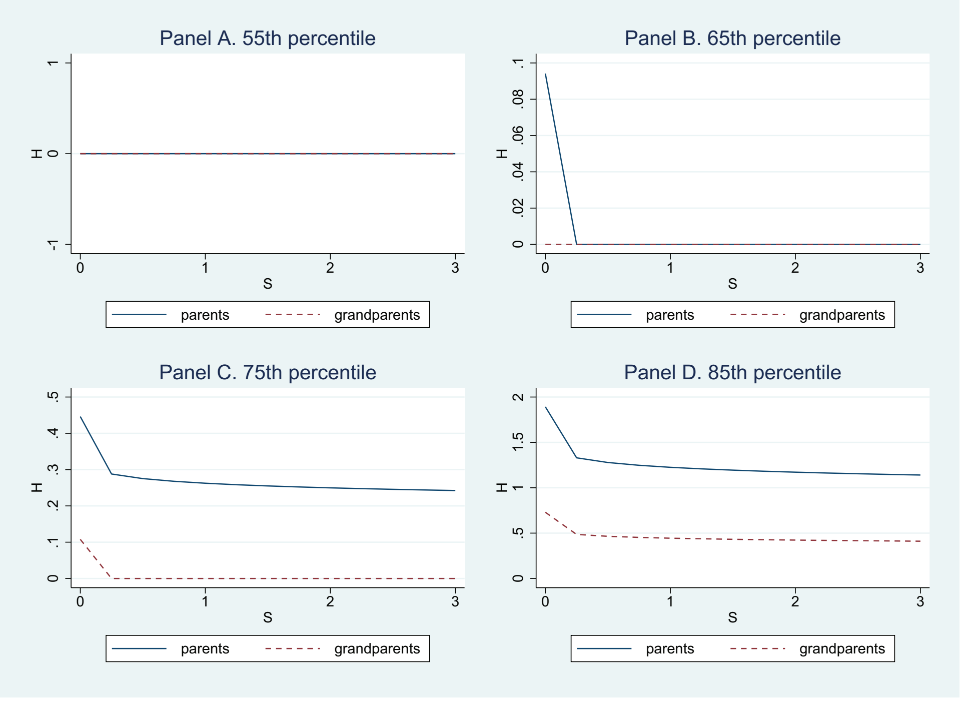


FIGURE 5

OPTIMAL CHOICE OF HOME-TUTORING TIME BY PERCENTILE RANK OF CAREGIVER'S VALUATION OF STUDENT ACHIEVEMENT

centile, both parents and grandparents choose $H^{**} = 0$ regardless of the values of S. We choose to start from the 55th percentile of the distribution of ω because the same results (i.e., $H^{**} = 0 \,\forall\, S$) hold for parents and grandparents with ω below the 55th percentile. At the 65th percentile (Panel B), parents start with a small positive H^{**} when $S = 0$ but quickly switch to the corner solution $H^{**} = 0$ when S increases, whereas grandparents always choose $H^{**} = 0$. At the 75th percentile (Panel C), parents and grandparents both start with positive home-tutoring time when $S = 0$. Although grandparents' home-tutoring time quickly switches to the corner solution $H^{**} = 0$ when S increases, parents continue choosing positive home-tutoring time even when school-tutoring time reaches the level of our intervention (i.e., three hours per week). At the 85th percentile (Panel D), parents and grandparents both start with positive home-tutoring time; when S increases from zero to three hours per week, both types of caregivers reduce their home-tutoring inputs but remain to choose positive H^{**}. In sum, the results in Figure 5 demonstrate a few patterns. First, an increase in school-tutoring time always (weakly) reduces home-tutoring time, suggesting that the two are substitutes. Second, at every percentile of ω demonstrated, parents choose (weakly) higher home-tutoring time than grandparents for any given level of school tutoring. Finally, the school-tutoring time has a (weakly) larger crowding-out effect on the home-tutoring time for non-LBC than LBC, and therefore, the difference in home-tutoring time between parents and grandparents decreases (weakly) with school-tutoring time. These three predictions are consistent with Propositions 1, 3, and 4, respectively.

Figure 6 further compares the MTE of school-tutoring time for children cared for by parents and grandparents. The MTE is defined as $\frac{dY}{dS}(= \frac{\partial Y}{\partial S} + \frac{\partial Y}{\partial H^{**}} + \frac{dH^{**}}{dS})$, which measures the net achievement gain from an extra unit of school tutoring after taking into account the caregiver's behavioral response in home-tutoring time. The four panels present the MTE corresponding to the 55th, 65th, 75th, and 85th percentile of the respective distribution of ω for parents and grandparents. In all panels, MTE is positive for all levels of school-tutoring time, sug-

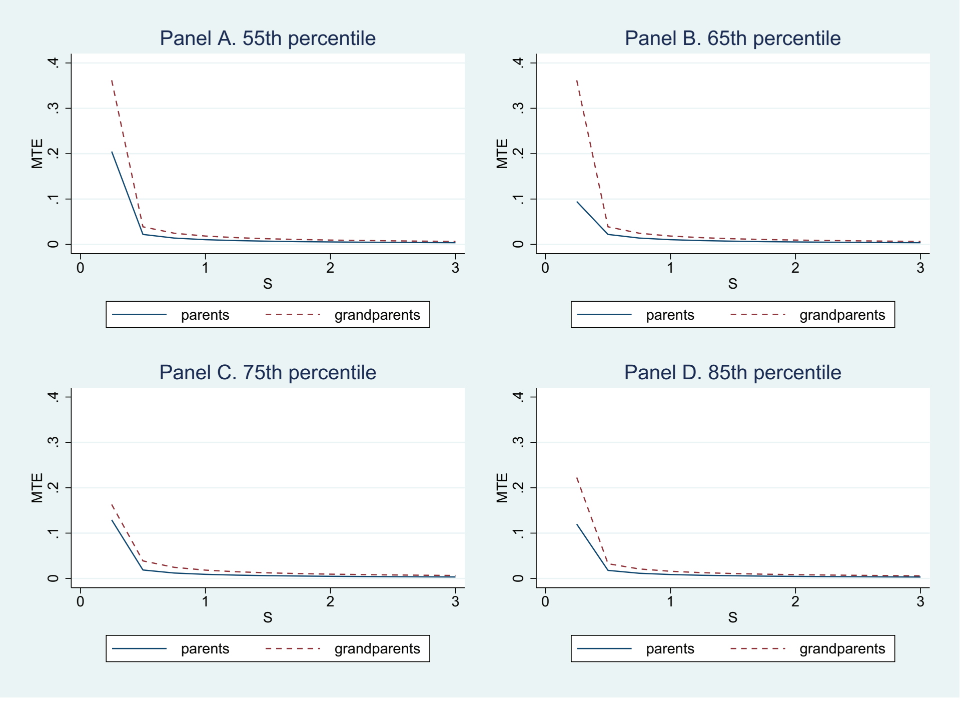


FIGURE 6

MARGINAL TREATMENT EFFECT OF SCHOOL TUTORING BY PERCENTILE RANK OF CAREGIVER'S VALUATION OF STUDENT ACHIEVEMENT

gesting that the effect of an increase in school tutoring on test scores always dominates the effect of the reduction (if any) in home tutoring. As we gradually increase school-tutoring time, the MTE declines for both groups of children due to the diminishing return to total tutoring. Moreover, the MTE of LBC is larger than that of non-LBC at all levels of school-tutoring time and percentiles of caregiver's valuation of student achievement. This suggests that a non-targeted, uniform school-tutoring program may not be the optimal policy. Instead, a targeted school-tutoring program providing more school tutoring to LBC can lead to efficiency gains.

Finally, we compare the current school-tutoring program, which provides equal school-tutoring time to LBC and non-LBC students, with an alternative program that optimally allocates different levels of school-tutoring time between LBC and non-LBC students. Suppose we have a 50–50 split of LBC and non-LBC, which is not far from our sample statistics of having 44% of children living without both parents. Figure 7 demonstrates our strategy to optimally split a total of s hours of school-tutoring time (ranging from 0 to 12 hours per week) between an LBC tutee and a non-LBC tutee. The horizontal axis is the school-tutoring time available for each tutee and the vertical axis is the group-specific MTE of school-tutoring time. The two solid lines present the MTE of school tutoring for LBC and non-LBC, respectively. Put in another way, for a given MTE level m in the vertical axis, the horizontal-axis values of the two solid lines correspond to $S_p^{-1}(m)$ and $S_g^{-1}(m)$, respectively, where $S_\tau^{-1}(m)$ for $\tau \in \{p, g\}$ denotes the inverse MTE function. The dashed line is the horizontal summation of the two solid lines, that is, $S_p^{-1}(m) + S_g^{-1}(m)$. For a given level of school-tutoring time s available for the pair of LBC and non-LBC, the interaction of the vertical line $S = s$ and the dashed line $S_p^{-1}(m) + S_g^{-1}(m)$ identifies the MTE under the optimal allocation $m^*(s)$. The interactions of the horizontal line $m = m^*(s)$ and the two solid lines identify $S_p^*(s)$ and $S_g^*(s)$, the optimal levels of school-tutoring time for LBC and non-LBC, respectively. An example demonstrated in Panel A of Figure 7 is that when the total school-tutoring time available for

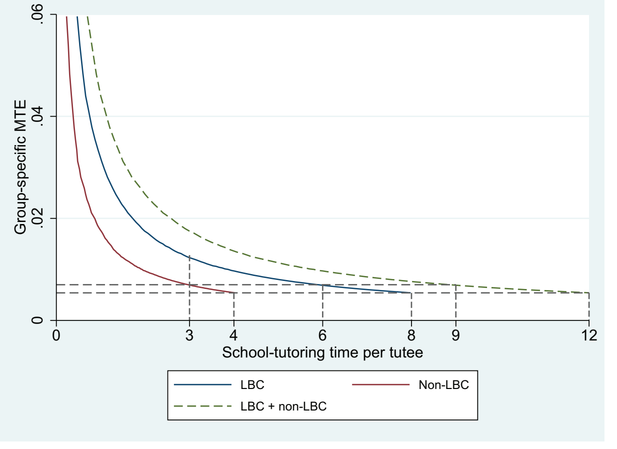


FIGURE 7

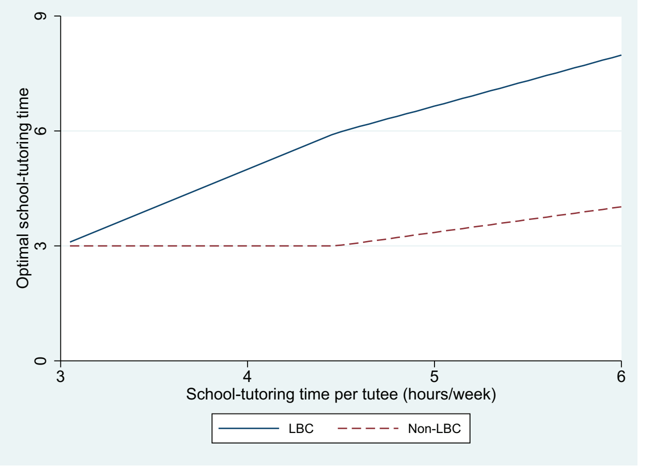
MARGINAL TREATMENT EFFECT: ALTERNATIVE SCHOOL TUTORING PROGRAMS

the pair of LBC and non-LBC is nine hours, the optimal split is to allocate six hours to the LBC and three hours to the non-LBC. Similarly, when the total school-tutoring time is 12 hours, we should assign eight hours to the LBC and four hours to the non-LBC to equalize their MTE.

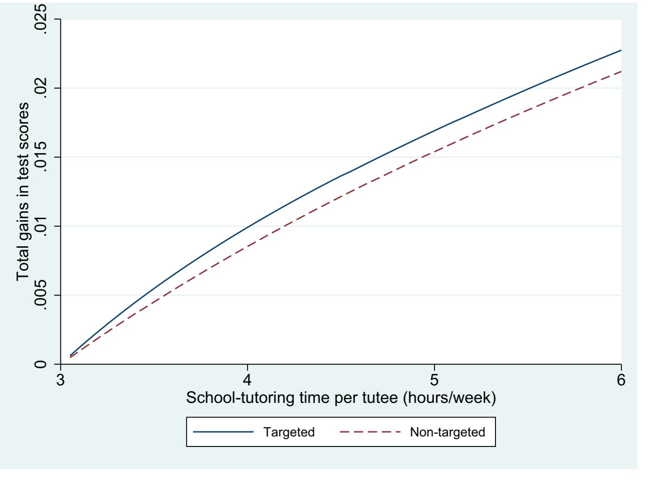
Now consider the case when we have additional school-tutoring resources on top of our existing experiment (three hours per week), how shall we allocate them between LBC and non-LBC? Panel A of Figure 8 presents the optimal assignment of school-tutoring time given the average school-tutoring time per tutee. It shows that when average school-tutoring time increases from 3 to 4.5 hours, all the school tutoring should be assigned to LBC. This is because the MTE at three to six hours per week for LBC is higher than the MTE at three hours per week for non-LBC, as shown in Figure 7. When the average school-tutoring time is higher than 4.5 hours, we should assign school tutoring to both LBC and non-LBC at a ratio of around 2.

Panel B of Figure 8 further presents the total gains in test scores at different levels of school-tutoring time compared to the baseline case when we assign three hours of school tutoring to both LBC and non-LBC. If we use the existing design (nontargeted program) that provides equal school tutoring to LBC and non-LBC, increasing school tutoring per week from three to four hours will lead to an increase in test scores by 0.0085σ. Using the targeted program that optimally distributes school-tutoring time between LBC and non-LBC (five hours to LBC and three hours to non-LBC, as shown in Panel A) will further increase the total gain by 16.5% (to 0.0099σ). Increasing school-tutoring time from three to six hours will increase test scores by 0.021σ in the nontargeted program, and the targeted program can further increase test scores by 9.5% (to 0.023σ). This counterfactual exercise suggests that the efficacy of the school-tutoring program can be improved by prioritizing the inputs to low-achieving students living with their grandparents. This is because they have higher tutoring efficiency and smaller household substitution in home-tutoring time.[30]

[30] If we consider an alternative baseline case when LBC and non-LBC have zero hours of school tutoring, the additional gain of moving from the nontargeted to the targeted program is much smaller. This is because the production function suggests that the MTE of a small amount of school tutoring is close to infinity for both LBC and non-LBC. Therefore, we should allocate school tutoring to both groups when they have no school tutoring. The optimal assignment will only matter when both groups already have a minimum level of school tutoring. Therefore, we choose to analyze the total gains for additional tutoring time on top of the three hours per week provided in the experiment.



Panel (a): Optimal assignment of school-tutoring time



Panel (b): Total gains in achievement

FIGURE 8

COUNTERFACTUAL: ALTERNATIVE SCHOOL TUTORING PROGRAMS

7. CONCLUSION AND DISCUSSION

Worldwide students' access to learning enrichment activities after school is highly dependent on family background, raising concerns over the implications of the advantages gained by children from higher-SES backgrounds through their families' time and money on educational inequality and social mobility. In rural China, these concerns are intensified by the presence of tens of millions of children left-behind by both parents who migrated in search of work in cities. These left-behind children are disadvantaged in after-school learning support received at home and also are academically lagging behind their peers living with parent(s), casting a suspicion that the negative achievement effects of being left-behind by both parents may work, at least in part, through the lack of after-school home-learning support (Zhang et al., 2014). Given the scale of parental absence in rural China, with one in six children living without both parents, there is a substantial need for compensatory programs to ameliorate the negative learning effects of parental absence. Nonetheless, despite the large and increasing policy interest in China and elsewhere for enhancing public roles in after-school learning support to reduce educational inequality, prior evidence of the effects of after-school programs on children's academic outcomes is limited and far from unified.

We conduct a randomized after-school tutoring experiment in a poor rural area of China in which high-achieving fourth and fifth graders provided high-dosage, one-on-one tutoring to low-achieving second and third graders. Prior to the experiment children who had been left-behind by both parents received far less home tutoring compared to children living with their parent(s). During the tutoring program, tutees living with parent(s) reported large significant reductions in home tutoring at both the extensive and intensive margins, whereas tutees without parents had much smaller, and often insignificant, reductions in home tutoring. We also find that the tutoring program significantly improved tutees' endline math scores, with the score gains being significantly larger for children without parents at home. In order to explain these empirical results, we develop and estimate a simple model of student achievement that integrates school and home tutoring into total tutoring to determine student achievement. Our structural estimation illustrates that the substitutability between home and school tutoring in achievement production (i.e., $\rho < \gamma$) and the higher returns to total tutoring for parents than grandparents (i.e., $\kappa_p > \kappa_g$) are the two underlying mechanisms driving the empirical findings of the experiment.

For an overall assessment of the program. in addition to estimates of the impacts, estimates of the costs are important. The direct costs of the intervention include overtime payments to the teachers supervising the tutorial sessions and stationery gifts to the voluntary tutors. Each teacher was paid RMB 300 per month over the eight-month intervention period and each tutor was given stationery gifts (distributed twice) equivalent to RMB 100. Thus, the total direct cost for the intervention is RMB 3,400 (USD 500) per tutorial session of 10 tutees, or RMB 340 (USD 50) per tutee. This is small, considering the achievement gains, especially those for the children living without parents.

Thus, our results demonstrate that supervised peer tutoring with professional-teacher support is a feasible and cost-effective remedial intervention for rural Chinese children lagging behind academically, particularly those left-behind by both parents. The opposing signs of the differences in the extent of substitution of home-tutoring inputs and test-score gains between children living with and without parents also yield *indirect* evidence that home inputs after school indeed matter for children's cognitive development, substantiating the concerns over the implications of differences in access to after-school learning opportunities by family background on child development and educational inequality. Moreover, an important policy implication of our results is that targeting public provision of after-school learning support to children from disadvantaged family backgrounds, such as children left-behind by their parents in rural China, is both an equitable and efficacious strategy: These children tend to lag behind in academic achievement but would experience larger achievement gains from after-school learning support owing to both greater tutoring efficacy and lesser substitution of school tutoring for household inputs. Last but not least, although we only implemented a particular form of after-school learning support (i.e., supervised peer tutoring), the insights that public provision of after-school learning support can particularly benefit children from disadvantaged family backgrounds and thus reduce educational inequality may be generalized for other forms of public support in after-school learning such as granting free access and user support to eLearning resources such as artificial intelligence learning services that more easily can be scaled up.

DATA AVAILABILITY STATEMENT. Data and codes used for producing the results of this article are publicly accessible from the openICPSR data repository at https://doi.org/10.3886/E193091V1.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table A1: Alternative Estimates from School-grade Fixed-effect Regressions

REFERENCES

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak, "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *Quarterly Journal of Economics* 126 (2011), 699–748.

Adhvaryu, A., A. Nyshadham, T. Molina, and J. Tamayo, "Helping Children Catch Up: Early Life Shocks and the Progresa Experiment," *Economic Journal*, forthcoming.

Aizer, A., and F. Cunha, "The Production of Human Capital: Endowments, Investments and Fertility," NBER Working Paper, Number 18429, 2012.

All-China Women's Federation, "A Study of the Situation of Left-Behind Children in Rural China," in Chinese, 2013, http://www.reformdata.org/2013/0510/22228.shtml (accessed 4 September 2023).

Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak, and C. R. Walters, "Inputs and Impacts in Charter Schools: KIPP Lynn," *American Economic Review* 100 (2010), 239–43.

———, ———, ———, ———, and ———, "Who Benefits from KIPP?," *Journal of Policy Analysis and Management* 31 (2012), 837–60.

———, P. A. Pathak, and C. R. Walters, "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics* 5 (2013), 1–27.

Athey, S., and G. Imbens, "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Science* 113(27) (2016), 7353–60.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, and M. Walton, "Teaching at the Right Level: Evidence from Randomized Evaluations in India," NBER Working Paper 22746, 2015.

Banerjee, A. V., ———, E. Duflo, R. Glennerster, and S. Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy* 2 (2010), 1–30.

———, S. Cole, ———, and L. Linden, "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics* 122 (2007), 1235–64.

Behrman, J. R., S. W. Parker, and P. E. Todd, "Incentives for Students and Parents, " in P. Glewwe, ed., *Education Policy in Developing Countries* (Chicago, IL: The University of Chicago Press, 2013), 137–92.

Bitler, M. P., H. W. Hoynes, and T. Domina, "Experimental Evidence on Distributional Effects of Head Start," NBER Working Paper, Number 20434, 2014.

Black, A. R., F. Doolittle, P. Zhu, R. Unterman, and J. B. Grossman, "The Evaluation of Enhanced Academic Instruction in After-School Programs: Findings After the First Year of Implementation," Technical Report, NCEE 2008-4021, 2008.

Boardman, A. E., and R. J. Murnane, "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement," *Sociology of Education* 52(2) (1979), 113–21.

Bray, M., and C. Lykins, *Shadow Education: Private Supplementary Tutoring and Its Implications for Policy Makers in Asia* (Manila: Asian Development Bank, 2012).

Cabezas, V., J. I. Cuesta, and F. A. Gallego, "Effects of Short-Term Tutoring on Cognitive and Non-Cognitive Skills: Evidence from a Randomized Evaluation in Chile," Unpublished manuscript, Pontificia Universidad Católica de Chile, Santiago, 2011.

Campbell, F., G. Conti, J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan, "Early Childhood Investments Substantially Boost Adult Health," *Science* 343 (2014), 1478–85.

Cascio, E. U., and D. W. Schanzenbach, "The Impacts of Expending Access to High Quality Preschool Education," *Brookings Papers on Economic Activity* Fall (2013), 161–93.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val, "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiment, with an Application to Immunization in India," NBER Working Paper, Number 24678, 2020.

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

Conn, K. M., "Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Impact Evaluations," *Review of Educational Research* 87 (2017), 863–98.

Cook, P. J., K. Dodge, G. Farkas, R. G. Fryer, J. Guryan, J. Ludwig, S. Mayer, H. Pollack, and L. Steinberg, "The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago," NBER Working Paper, Number 19862, 2014.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg, "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance," *Journal of Political Economy* 126 (2018), 2356–409.

Cunha, F., and J. Heckman, "The Technology of Skill Formation," *American Economic Review* 97 (2007), 31–47.

———, J. J. Heckman, and S. M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica* 78 (2010), 883–931.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman, "School Inputs, Household Substitution, and Test Scores," *American Economic Journal: Applied Economics* 5 (2013), 29–57.

Davis, J. M., and S. B. Heller, "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs," *Review of Economics and Statistics* 102(4) (2020), 664–77.

Dobbie, W., and R. G. Fryer, "Getting Beneath the Veil of Effective Schools: Evidence from New York City," *American Economic Journal: Applied Economics* 5 (2013), 28–60.

Duflo, E., P. Dupas, and M. Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review* 101 (2011), 1739–74.

Duncan, G. J., and R. J. Murnane, "Rising Inequality in Family Incomes and Children's Educational Outcomes," *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2 (2016), 142–58.

Fryer, R. G., "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *Quarterly Journal of Economics* 129 (2014), 1355–407.

Gelber, A., and A. Isen, "Children's Schooling and Parents' Behavior: Evidence from the Head Start Impact Study," *Journal of Public Economics* 101 (2013), 25–38.

Goff, L., O. Malamud, C. Pop-Eleches, and M. Urquiola, "Interactions Between Family and School Environments: Access to Abortion and Selective Schools," *Journal of Human Resources*, forthcoming.

Guryan, J., E. Hurst, and M. Kearney, "Parental Education and Parental Time with Children," *Journal of Economic Perspectives* 22 (2008), 23–46.

Hanushek, E., "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources* 14 (1979), 351–88.

Havnes, T., and M. Mogstad, "Is Universal Child Care Leveling the Playing Field?," *Journal of Public Economics* 127 (2015), 100–114.

Heckman, J., S. H. Moon, R. Pinto, P. Savelyev, and A. Yavitz, "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program," *Quantitative Economics* 1 (2010), 1–46.

———, R. Pinto, and P. Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," 103 (2013), 2052–86.

Heckman, J. J., "The Economics, Technology, and Neuroscience of Human Capability Formation," *Proceedings of the National Academy of Sciences* 104 (2007), 13250–55.

Helpman, E., M. Melitz, and Y. Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics* 123 (2008), 441–87.

James-Burdumy, S., M. Dynarski, and J. Deke, "When Elementary Schools Stay Open Late: Results from the National Evaluation of the 21st Century Community Learning Centers Program," *Educational Evaluation and Policy Analysis* 29 (2007), 296–318.

Kane, T. J., D. F. McCaffrey, T. Miller, and D. O. Staiger, "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment," Research Paper. MET Project. Bill & Melinda Gates Foundation (Citeseer, 2013).

———, and D. O. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," Technical Report, National Bureau of Economic Research, 2008.

Kim, S., and J.-H. Lee, "Private Tutoring and Demand for Education in South Korea," *Economic Development and Cultural Change* 58 (2010), 259–96.

Kottelenberg, M. J., and S. F. Lehrer, "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Child Care," *Journal of Labor Economics* 35 (2017), 609–53.

LEE, D. S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies* 76 (2009), 1071–102.

LI, H., P. LOYALKA, S. ROZELLE, B. WU, and J. XIE, "Unequal Access to College in China: How Far Have Poor, Rural Students Been Left Behind?," *The China Quarterly* 221 (2015), 185–207.

LI, T., L. HAN, L. ZHANG, and S. ROZELLE, "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools," *Journal of Public Economics* 111 (2014), 29–45.

LUCAS, A. M., and I. M. MBITI, "Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya," *American Economic Journal: Applied Economics* 4 (2012), 226–53.

MATSUOKA, R., "School Socioeconomic Composition Effect on Shadow Education Participation: Evidence from Japan," *British Journal of Sociology of Education* 36 (2015), 270–90.

MBITI, I., K. MURALIDHARAN, M. ROMERO, Y. SCHIPPER, C. MANDA, and R. RAJANI, "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania," *Quarterly Journal of Economics* 134 (2019), 1627–73.

MILLS, J. N., and P. J. WOLF, "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement After 2 Years," *Educational Evaluation and Policy Analysis* 39 (2017), 464–84.

MURALIDHARAN, K., A. SINGH, and A. J. GANIMIAN, "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India," *American Economic Review* 109 (April 2019), 1426–60.

NICKOW, A., P. OREOPOULOS, and V. QUAN, "The Impressive Effects of Tutoring on Prek-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence," NBER Working Paper, Number 27476, 2020.

OAXACA, R., and M. RANSOM, "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics* 61 (1994), 5–21.

POP-ELECHES, C., and M. URQUIOLA, "Going to a Better School: Effects and Behavioral Responses," *American Economic Review* 103 (2013), 1289–324.

RAMEY, G., and V. A. RAMEY, "The Rug Rat Race," *Brookings Papers on Economic Activity* 2010 (2010), 129–76.

ROSSIN-SLATER, M., and M. WÜST, "What Is the Added Value of Preschool? Long-Term Impacts and Interactions with a Health Intervention," *American Economic Journal: Applied Economics* 12 (2020), 255–86.

SAFARZYŃSKA, K., "Socio-Economic Determinants of Demand for Private Tutoring," *European Sociological Review* 29 (2013), 139–54.

STATE COUNCIL OF CHINA, "Opinions on Regulating the Development of Off-Campus Training Institutions," General Office of the State Council Document No. 80 (in Chinese), 2018.

SUMMERS, A. A., and B. L. WOLFE, "Do Schools Make a Difference?," *The American Economic Review* 67 (1977), 639–52.

TODD, P. E., and K. WOLPIN, "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps," *Journal of Human Capital* 1 (2007), 91–136.

———, and K. I. WOLPIN, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal* 113 (2003), F3–F33.

U.S. DEPARTMENT OF EDUCATION, *21st Century Community Learning Centers (21st CCLC) Analytic Support for Evaluation and Program Monitoring: An Overview of the 21st CCLC Performance Data: 2013-14* (Washington, DC: U.S. Department of Education, 2015).

WAGER, S., and S. ATHEY, "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of American Statistical Association* 113 (2018), 1228–42.

WEI, Y., *Report on Household Education Expenditure in China (in Chinese)* (China Institute of Education Finance Research, 2019), http://ciefr.pku.edu.cn/cbw/kyjb/2018/03/kyjb_5257.shtml(accessed 18 April 2019).

WEISS, H. B., P. M. LITTLE, S. M. BOUFFARD, S. N. DESCHENES, and H. J. MALONE, "The Federal Role in Out-of-School Learning: After-School, Summer Learning, and Family Involvement as Critical Learning Supports," A Research Review Paper and Recommendations from Harvard Family Research Project, 2009.

WEISS, M. J., A. RATLEDGE, C. SOMMO and H. GUPTA, "Supporting Community College Students from Start to Degree Completion: Long-Term Evidence from a Randomized Trial of CUNY's ASAP," *American Economic Journal: Applied Economics* 11 (2019), 253–97.

YUAN, C., and L. ZHANG, "Public Education Spending and Private Substitution in Urban China," *Journal of Development Economics* 115 (2015), 124–39.

ZHANG, D., X. LI, and J. XUE, "Education Inequality between Rural and Urban Areas of the People's Republic of China, Migrants' Children Education, and Some Implications," *Asian Development Review* 32 (2015), 196–224.

ZHANG, H., "Identification of Treatment Effects under Imperfect Matching with an Application to Chinese Elite Schools," *Journal of Public Economics* 142 (2016), 56–82.

———, J. R. Behrman, C. S. Fan, X. Wei, and J. Zhang, "Does Parental Absence Reduce Cognitive Achievements? Evidence from Rural China," *Journal of Development Economics* 111 (2014), 181–95.

Zhang, W., "The Demand for Shadow Education in China: Mainstream Teachers and Power Relations," *Asia Pacific Journal of Education* 34 (2014), 436–54.

Zhang, Y., and Y. Xie, "Family Background, Private Tutoring, and Children's Educational Performance in Contemporary China," *Chinese Sociological Review* 48 (2016), 64–82.

Zhao, G., J. Ye, Z. Li, and S. Xue, "How and Why Do Chinese Urban Students Outperform Their Rural Counterparts?," *China Economic Review* 45 (2017), 103–23.