PENN INSTITUTE for ECONOMIC RESEARCH
UNIVERSITY of PENNSYLVANIA

The Ronald O. Perelman Center for Political Science and Economics (PCPSE)
133 South 36th Street
Philadelphia, PA 19104-6297

pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 20-004

After-School Tutoring, Household Substitution and Student Achievement: Experimental Evidence from Rural China

JERE R. BEHRMAN
University of Pennsylvania

C. SIMON FAN
Lingnan University

XIANGDONG WEI
Lingnan University

HONGLIANG ZHANG
Hong Kong Baptist University

JUNSEN ZHANG
The Chinese University of Hong Kong

January 20, 2020

https://ssrn.com/abstract=3523203

After-School Tutoring, Household Substitution and Student Achievement: Experimental Evidence from Rural China*

Jere R. Behrman[1], C. Simon Fan[2], Xiangdong Wei[2], Hongliang Zhang[3], and Junsen Zhang[4]

[1]Departments of Economics and Sociology, University of Pennsylvania, United States

[2]Department of Economics, Lingnan University, Tuen Mun, Hong Kong

[3]Department of Economics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[4]Department of Economics, The Chinese University of Hong Kong, Shatin, Hong Kong

*This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0005294. We thank Joshua Angrist, 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 acknowledge funding support provided by the National Natural Science Foundation of China (No. 71173178), Hong Kong Research Grants Council General Research Fund (No. 458910), and the Chinese University of Hong Kong Research Committee Group Research Scheme. We are also grateful to the Educational Bureau of the Longhui County in China for administrative support, and Chuantao Cui, Bei Luo, Amy Ru Chien Tseng, Baojun Wang, Jia Wu, and Xiang Zhou for excellent research assistance.

Abstract

Worldwide children's access to after-school learning activities is highly dependent on family backgrounds. Concern over the implications of such activities for child development and educational inequality has led to a global rise of public provision of after-school learning support. However little is known about interactions of public after-school activities and household investments in children's learning. This paper contributes to the literature on the effects of public inputs on household inputs and student achievement in after-school settings. We build a model that integrates public and private inputs to produce student achievement through two competing mechanisms – diminishing returns to total inputs and complementarity between public and private inputs. When diminishing returns dominate complementarity, the model predicts the substitution away of private inputs due to increases in public inputs for all households, although the extent of crowding-out is smaller and therefore the test score gains are larger for children from disadvantaged family backgrounds facing higher costs of private inputs. We implement a randomized controlled after-school tutoring experiment in rural China where many children are left-behind by both parents and cared for by grandparents. During the program, tutees living with parents reported large and significant reductions in the amount of tutoring received at home, whereas tutees living apart from both parents reported much smaller, and often insignificant, reductions. We find that tutees' math scores improved significantly, and more for children living without parents, although there is no evidence for improvement in tutees' endline reading scores.

1 Introduction

Today, in much of the world, students engage in academically-focused after-school learning activities, taking the form of either family direct involvement or supplementary educational services. However, almost everywhere, children's access to after-school learning opportunities is highly dependent on family backgrounds (e.g., Weiss et al., 2009; Bray and Lykins, 2012). For example, in the U.S., Guryan, Hurst and Kearney (2008) show that college-educated mothers spend on average 16.5 hours per week in child care compared to only 12.1 hours for high school dropouts[1] and Duncan and Murnane (2016) find that top-quintile families spend seven times more than bottom-quintile families on child-enrichment activities. 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). The advantage gained by children with higher SES through their families' private resource investments, in terms of both time and money, is a source of educational inequality that reduces social mobility (Park et al., 2016).

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 U.S., the 21^{st} Century Community Learning Centers (21^{st} CCLC) program was authorized under the No Child Left Behind Act (NCLB) as an after-school program model to provide academic enrichment and services during non-school hours to help students attending high-poverty, underperforming schools meet federal and state standards in core academic subjects.[2] 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 consecu-

[1] In another study, Ramey and Ramey (2010) show that the most important uses of the extra time spent by college-educated parents are teaching children and organizing and attending extracurricular activities.

[2] 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, Dynarski and Deke, 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 2013-2014 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).

tive years. In South Korea, where the average family spends nearly 20% of its income on private tutoring, the Ministry of Education has offered extra-hour, in-campus cramming sessions within public schools and television-based tutorial lessons via the nation's Educational Broadcasting System[3] to provide an option, particularly for students from poorer families, other than private after-school tutoring (Chandler, 2011). In China, which has over 400,000 after-school educational institutions and an estimated after-school market size of RMB 800 billion (USD 115 billion),[4] the State Council recently issued an official document to launch a national campaign to tighten scrutiny of private after-school institutions and enhance the public role in providing after-school services (State Council of China, 2018).

Yet despite the enormous policy interest, evidence of the effects of public after-school programs on children's academic outcomes is limited and far from unified. Whereas Meyer and Van Klaveren (2013), Miller and Connolly (2013), and Gaastra (2016) find no evidence of academic benefits from after-school program participation in the Netherlands, Northern Ireland, and San Diego, respectively, significant positive effects are reported by Banerjee et al. (2010) in India, Zimmer, Hamilton and Christina (2010) in Pittsburg, and Cook et al. (2014) in Chicago. Note that these studies estimate the gross or total policy effects of programs that combine the ceteris paribus effects holding other inputs constant and the indirect effects through changes in private inputs after household re-optimization (Todd and Wolpin, 2003). While some differences in the estimates are undoubtedly due to differences in program features, differences in household behavioral responses in private inputs also may be important sources of variation of the estimated gross policy effects. Although having received little attention in the aforementioned studies of public after-school programs, the empirical relationship between private household inputs and public inputs has been investigated in other contexts outside the after-school settings, albeit with mixed results even in terms of the direction of household behavioral responses. For example, whereas Pop-Eleches and Urquiola (2013) and Yuan and Zhang (2015) find evidence of substitution between parental

[3]The Ministry of Education of South Korea mandated that 70% of questions on the national college entrance exam be based on lessons carried on the government-funded Educational Broadcasting System, thus creating a strong incentive for students to tune into this public program.

[4]These numbers are from two China Daily articles: `http://www.chinadaily.com.cn/a/201811/22/WS5bf60804a310eff30328a54e.html` (accessed 18 April 2019) and `http://www.chinadaily.com.cn/business/2016-12/29/content_27810429.htm` (accessed 18 April 2019).

inputs and school resources in Romania and China, Gelber and Isen (2013) show that parental involvement with children increases in response to access to the Head Start Program in the U.S. Moreover, in a study of both parental responses and achievement impacts of increases in school grants in India and Zambia, Das et al. (2013) find that only unanticipated increases in school resources – which are not offset by household own educational spending – have positive effects on test scores, whereas anticipated increases in school resources crowd out private household educational expenditures and generate no test score gains. Given the great heterogeneity in household behavioral responses found across programs and contexts and their relevance for the overall achievement effect that Das et al. (2013) demonstrate, it seems that household behavioral responses probably play significant roles in determining the overall effectiveness of public interventions in after-school settings.

In this paper, we first develop a simple model of student achievement production that features (i) diminishing returns to aggregate total tutoring received after school and (ii) complementarity between school-based, public tutoring inputs and home-based, private tutoring inputs in generating total tutoring. This model predicts that increases in public tutoring inputs crowd-out private tutoring inputs when diminishing returns dominate complementarity but encourage private tutoring inputs otherwise. Moreover, if there exists household heterogeneity in private tutoring costs,[5] the model also predicts that behavioral responses and treatment effects differ across households. In particular, in the case in which increases in public tutoring inputs crowd-out private tutoring inputs (i.e., diminishing returns dominate complementarity), the model predicts greater substitution away from private tutoring inputs and therefore smaller test score gains for children from households relatively more cost-effective in private tutoring provision.

We then proceed to test the model predictions with regard to increases in public tutoring inputs using a randomized after-school tutoring experiment in which high-achieving 4^{th} and 5^{th} graders provided high-dosage, one-on-one tutoring to low-achieving 2^{nd} and 3^{rd} graders. We take advantage of the widespread "left-behind children" phenomenon in rural China where many children are living apart from both parents who have migrated to work in cities, and implement the experiment in a rural Chinese county with a high prevalence of left-behind

[5] That is, households differ in their costs of providing the same effective amount of private tutoring inputs.

children to assess household behavioral response and treatment effect heterogeneity between children with and without parents at home. Prior to the experiment children who had been left-behind by both parents – most of whom were cared for by grandparents – received far less home tutoring compared to children living with their parent(s).[6] During participation in the tutoring program, tutees living with parent(s) reported large and significant reductions in home tutoring at both the extensive and intensive margins, whereas tutees living apart from both parents 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 children without both parents at home. However, we find little evidence that the program increased tutees' endline reading scores. The disparity in the achievement effects between math and reading is not a complete surprise, as previous work on the efficacy of other educational interventions across subject areas also tend to find these interventions more effective for improving math scores than reading scores (e.g., Abdulkadiroğlu et al., 2011; Angrist et al., 2010; Black et al., 2008; Dobbie and Fryer, 2013).

This paper has several important contributions to the literature on the effects of public schooling inputs on private household inputs and student achievement. First, we develop a model that integrates public and private inputs to produce student achievement through two competing mechanisms – diminishing returns to aggregate tutoring inputs and complementarity between public and private inputs in generating the aggregate tutoring inputs. This model reconciles the divergent evidence of household responses to public schooling inputs found in different settings and contexts in prior research. Second, we extend the empirical investigation of this strand of literature to household behavioral responses and treatment effects of increases in public inputs in after-school learning support, which may cause greater substitution of private inputs compared to other forms of public inputs but has nonetheless been largely overlooked in prior research. Third, we exploit the "left-behind children" phenomenon in rural China to examine heterogeneity in household behavioral responses and treatment effects of access to an after-school tutoring intervention. By linking the difference in the extent of substitution of home-tutoring inputs between children with and without parents at home to the difference in their score gains, we are able to gauge more directly

[6] Throughout the paper, "living with parent(s)" means living with either or both parents.

the implications of household behavioral responses on the total effect of the intervention. Fourth, 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 selected as tutees to participate in the after-school tutoring program, the empirical analysis of the paper is also related to the literature on remedial educational interventions targeted to children lagging behind academically (e.g., Banerjee et al., 2007, 2010; Duflo, Dupas and Kremer, 2011). In particular, a strand of this research has employed randomized controlled trials (RCTs) to evaluate the effectiveness of high-dosage, small-group or one-on-one tutoring outside the regular school hours. Cabezas, Cuesta and Gallego (2011) implement a RCT of a three-month small-group tutoring program to low-achieving 4^{th} graders in Chile using college student volunteers as tutors, and find that the program has significant positive effects only for a subsample of students coming from low-performing and poor schools. Cook et al. (2014) randomly assign 106 disadvantaged male high school students in Chicago who had lagged behind in both academic and non-academic achievements to an intervention with both high-dosage, individualized small-group math tutoring and behavioral therapy, and find that the intervention increased participants' math scores by 0.65 standard deviations (hereafter σ) and expected graduation rates by 14 percentage points. Weiss et al. (2019) randomly assign 896 low-income community college students with remedial education needs in New York to experimental and control groups, and find that providing the students in the experimental group with tutoring outside of class, career guidance, and financial support significantly raises their college graduation rates. Li et al. (2014) randomly pair high- and low-achieving classmates as benchmates in primary schools in urban China that enroll exclusively children of migrant workers without local hukou registration and offer the pairs group incentives for the lower-achiever's score improvement. They find that the intervention increases lower achievers' test scores by 0.27σ without harming high achievers, although this overall effect combines influences of group incentives and peer interactions, the latter of which may also include peer tutoring outside regular school hours. Our study contributes to this strand of research in three aspects. First, we highlight both theoretically and empirically the important roles of household substitution behaviors in determining the effectiveness of remedial educational programs. Second, we demonstrate that remedial educational interventions are most effective for children who come from disadvantaged family backgrounds and have less home-learning support. Third, our results also imply that remedial educational programs offering learning inputs in settings that are subject to lesser substitution of household inputs will generate larger effects.

Finally, this paper further adds to an emerging literature demonstrating that children from disadvantaged backgrounds benefit the most from certain public educational interventions, such as universal child care programs (e.g., Cascio and Schanzenbach, 2013; Bitler, Hoynes and Domina, 2014; Havnes and Mogstad, 2015) and charter schools or traditional public schools that inject successful practices of charter 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 children from disadvantaged family backgrounds benefit the most from the universal child care program, and Angrist, Pathak and Walters (2013) show that Massachusetts charter schools adhering to the "No Excuse" model[7] are most effective for poor nonwhites and low-baseline achievers. Our findings of larger math 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 learning of children who have been lagging behind academically but also acts as an important equalizer for reducing inequality in child development by family backgrounds.

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.

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

2 Background

With an estimated market size of RMB 800 billion (or USD 115 billion), China now dominates the world's shadow educational industry and accounts for eight of the top 15 listed companies globally.[8] According to the first large-scale national survey on household educational expenditure carried out in 2017 by the China Institute of Education Finance Research (CIEFR), 37.8% of the country's elementary and secondary school students participated in academically-focused after-school programs and spent on average RMB 5,021 (USD 733) per pupil per year (Wei, 2017). However, similar to elsewhere in the world (e.g., Bray, 2007), shadow education is also predominately an urban phenomenon in China. 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.[9] Accompanying this difference in shadow education use is an even more striking urban-rural disparity in children's educational attainments. Using a nationwide dataset of college entrance examinations and admissions, Li et al. (2015) show that rural youth from poor counties[10] are eight times less likely (2% vs. 16%) than urban youth to attend a four-year college.[11] 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 calling for public elementary and secondary schools to enhance their roles in offering after-school services. 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

[8] `https://www.scmp.com/business/companies/article/2159159/mainland-chinas-after-school-tutoring-industry-ripe-consolidation` (last accessed on 19 April 2019).

[9] Wei (2017) also shows that even among households who had used academically-focused after-school tutoring programs, urban households on average spent 3.6 times that spent by their rural counterparts.

[10] Li et al. (2015) define "poor counties" as 592 nationally-designated poor counties identified by the Chinese government in 2003, one of which includes our experimental site.

[11] Zhang, Li and Xue (2015) and Zhao et al. (2017) 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.

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 by both parents (All-China Women's Federation, 2014). 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).[12] Moreover, we also find that absence of both parents is associated with significantly lower household educational inputs, mainly 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 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 Model

To guide our empirical analysis, we present a simple model that integrates public and private tutoring inputs to determine student achievement. Since we randomly assigned students from the same school-grades into the experimental and control groups, all school-level factors

[12] Specifically, Zhang et al. (2014) apply dynamic panel methods to longitudinal data on parental migration and children's test scores to evaluate te respective effects of being left-behind by one or both parents on children's cognitive achievements. Their results indicate that being left-behind by both parents reduces children's contemporary achievements by 5.4 percentile points for math and 5.1 percentile points for reading.

and the predetermined individual- or family-level factors determining student achievement are balanced between the treatment and control groups in our empirical estimates. Thus, to simplify, our model abstracts from other aspects of achievement determination such as school quality, ability endowments, and dynamic complementarities considered in prior research(e.g., Todd and Wolpin, 2003; Cunha and Heckman, 2007) and focuses on *contemporary* tutoring inputs only. In the model, we assume that student achievement (Y) depends only on total tutoring received (T), which itself is a CES aggregation of tutoring received in school (S) and tutoring received at home (H).

$$Y = A \cdot T^{\rho} = A \cdot \left[\theta S^{\gamma} + (1-\theta) H^{\gamma}\right]^{\frac{\rho}{\gamma}}, \tag{1}$$

where $A > 0$ represents the efficiency of total tutoring T in promoting achievement; ρ is the degree of scale of the achievement production function – we focus on diminishing returns to scale $\rho \in (0,1)$; $\theta \in (0,1)$ is the share parameter of school-tutoring inputs S; and $\gamma \in (-\infty, 1)$ with $\frac{1}{1-\gamma}$ corresponding to the elasticity of substitution between S and H.[13]

Within the CES construction, S and H are always *direct complements* in producing T, i.e.,

$$\frac{\partial^2 T}{\partial S \partial H} = (1-\gamma)\theta(1-\theta)\theta \left[\theta S^{\gamma} + (1-\theta) H^{\gamma}\right]^{\frac{1}{\gamma}-2} S^{\gamma-1} H^{\gamma-1} > 0.$$

However, because of diminishing returns to T in achievement production, S and H are not necessarily *direct complements* in producing achievement Y. To see this, take the cross derivative of Y with respect to S and H in Equation (1):

$$\frac{\partial^2 Y}{\partial S \partial H} = A\rho(\rho-\gamma)\theta(1-\theta) \left[\theta S^{\gamma} + (1-\theta) H^{\gamma}\right]^{\frac{\rho}{\gamma}-2} S^{\gamma-1} H^{\gamma-1}. \tag{2}$$

Since all terms in Equation (2) 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 to T in producing Y, whereas smaller γ indicates greater complementary between H and

[13]For ease of illustration and interpretation, we exclude from Equation (1) the special case $\gamma = 1$, under which S and H are perfect substitutes (i.e., no complementarity between them) in producing T. Note that all of the conclusions of the model hold for this special case as the condition $\rho < \gamma$, is satisfied if $0 < \rho < 1$ and $\gamma = 1$.

S in their aggregation to T.[14] Thus, S and H are *direct substitutes* in producing achievement ($\frac{\partial^2 Y}{\partial S \partial H} < 0$) when diminishing returns dominate complementary ($\rho < \gamma$), and *direct complements* in producing achievement if otherwise.

We next consider the household's optimization problem. Given the exogenously determined level of school-tutoring inputs S, households optimally choose own tutoring inputs H to maximize the child's achievement net of the cost of home-tutoring efforts. Note that in Equation (1), T , S , and H are all measured in efficiency units in producing achievement.[15] Regarding home-tutoring inputs H, for ease of illustration, we assume a constant cost per effective unit of home-tutoring inputs ω such that a household incurs a total cost ωH for choosing (effective) home-tutoring inputs H. However, we allow ω to be heterogenous across households due to their differences in tutoring effectiveness and opportunity costs. Although parents may have higher efficiencies in both tutoring and market/home production than grandparents, we assume here the dominance of the former over the latter such that parents have relative advantages in tutoring and therefore a lower ω for each (effective) unit of home-tutoring inputs than grandparents.

The household's objective function is thus $A[\theta S^\gamma + (1-\theta)H^\gamma]^{\frac{\rho}{\gamma}} - \omega H$, additively separable between child achievement and cost of home-tutoring efforts. Given the exogenously determined values of ω and S, the unconstrained optimal level of home-tutoring inputs H^* is determined by

$$A\rho(1-\theta)\left[\theta S^\gamma + (1-\theta){H^*}^\gamma\right]^{\frac{\rho}{\gamma}-1} {H^*}^{\gamma-1} = \omega. \tag{3}$$

As shown in Appendix A, applying the implicit function theorem to Equation (3) yields the following:

$$\frac{dH^*}{d\omega} = -\frac{1}{A\rho(1-\theta)\left[\theta S^\gamma + (1-\theta){H^*}^\gamma\right]^{\frac{\rho}{\gamma}-2} {H^*}^{\gamma-2}\left[(1-\rho)(1-\theta){H^*}^\gamma + (1-\gamma)\theta S^\gamma\right]}, \tag{4}$$

and

$$\frac{dH^*}{dS} = \frac{(\rho-\gamma)\theta S^{\gamma-1}}{(1-\rho)(1-\theta){H^*}^{\gamma-1} + (1-\gamma)\theta S^\gamma {H^*}^{-1}}. \tag{5}$$

[14] In the extreme case in which $\gamma = -\infty$, S and H are perfect complements in producing T such that $T = min\{\theta S, (1-\theta)H\}$.

[15] Taking home-tutoring inputs H as an example, various forms of such inputs differing in nature, amount, and efficacy are converted and normalized into efficiency units in producing achievement.

Under the parameter space considered here (i.e., $0 < \theta < 1, 0 < \rho < 1$ and $\gamma < 1$), all terms in the denominator of Equation (4) are positive and therefore $\frac{dH^*}{d\omega}$ must be negative. Thus, given our assumption that parents have relative advantages in home tutoring (i.e., a lower ω) than grandparents, Equation (4) yields the following hypothesis.

Hypothesis 1 *At the same level of school-tutoring inputs, children cared for by parents receive a higher level of home-tutoring inputs than those cared for by grandparents.*

Equation (5) relates household's optimal level of home-tutoring inputs H^* to tutoring received in school S. Since all terms in this equation are positive except possibly $\rho - \gamma$, $\frac{dH^*}{dS}$ has the same sign as $\rho - \gamma$. In particular, when S and H are *direct substitutes* in producing achievement (i.e., $\rho < \gamma$), $\frac{dH^*}{dS} < 0$ in Equation (5), which yields the following hypothesis regarding household responses in own tutoring inputs to increases in tutoring received in school.

Hypothesis 2 *Household's optimal home-tutoring inputs decrease with school-tutoring inputs.*

Thus, under the condition $\rho < \gamma$, although the direct effects of increased school-tutoring inputs on student achievement ($\frac{\partial Y}{\partial S}$) are positive, the indirect effects through household's endogenous adjustment in home-tutoring inputs ($\frac{\partial Y}{\partial H} \cdot \frac{dH^*}{dS}$) are negative. Nonetheless, as we show formally in Appendix B, the positive direct effects dominate the negative indirect effects such that the overall effects of increased school-tutoring inputs on student achievement remains positive ($\frac{dY}{dS} = \frac{\partial Y}{\partial S} + \frac{\partial Y}{\partial H} \cdot \frac{dH^*}{dS}$) despite household's substitution of home-tutoring inputs.

Hypothesis 3 *The positive direct effects of increased school-tutoring inputs dominates the negative indirect effects through household's optimal substitution in home-tutoring inputs, resulting in an overall increase in student achievement.*

Finally, we consider how the adjustments in the optimal level of home-tutoring inputs ($\frac{dH^*}{dS}$) and the total achievement effect of increased school-tutoring inputs ($\frac{dY}{dS}$) differ across households with different unit costs of home-tutoring inputs ω, Under the condition $\rho < \gamma$, we derive formally in Appendix C that the two cross-derivatives $\frac{d^2H^*}{dSd\omega}$ and $\frac{d^2Y}{dSd\omega}$ are both positive. With $\frac{dH^*}{dS} < 0$ (*Hypothesis 2*), $\frac{d^2H^*}{dSd\omega} > 0$ indicates that when school-tutoring inputs increase, households with higher ω reduce their home-tutoring inputs less than households with lower ω. Analogously, with $\frac{dY}{dS} > 0$(*Hypothesis 3*), $\frac{d^2Y}{dSd\omega} > 0$ indicates that the gross achievement effect of increased school tutoring is greater for children from households facing higher ω. Given our assumption that parents have relative advantages in home tutoring (i.e., a lower ω) than grandparents, these results yield the following hypothesis.

Hypothesis 4 *When school-tutoring inputs increase,*
(a) Grandparents reduce their own tutoring inputs less than parents.
(b) Children cared for by grandparents experience greater achievement gains than children cared for by parents.

Figure 1 summarizes the predictions of our theoretical analysis regarding households' optimal choices of home-tutoring inputs when diminishing returns in the achievement production dominate the complementary between school- and home-tutoring inputs in the aggregation of total tutoring inputs (i.e., $\rho < \gamma$). It compares two types of households with the children cared for by parents and grandparents, and assumes a lower unit cost of effective home-tutoring inputs for parents than grandparents, $\omega_p < \omega_g$, with the subscript p and g denoting parents and grandparents, respectively. First, with an initial level of school-tutoring inputs S, parents choose a higher optimal level of home-tutoring inputs than grandparents,$H_p^* > H_g^*$ (*Hypothesis 1*). Second, when the level of school-tutoring inputs increases from S to S', both parents and grandparents reduce their level of home-tutoring inputs, $H_p^{*\prime} < H_p^*$ and $H_g^{*\prime} < H_g^*$ (*Hypothesis 2*). Third, the extent of the reduction in home-tutoring inputs is greater for parents than grandparents, $H_p^* - H_p^{*'} > H_g^* - H_g^{*'}$ (*Hypothesis 4a*). Although not demonstrated in Figure 1, *Hypothesis 3* and *Hypothesis 4b* further predict the total achievement effect of increased school-tutoring inputs after taking into account household behavioral responses: while all children benefit from increased school-tutoring inputs regardless of who their primary caregivers are (*Hypothesis 3*), the extent of the achievement gains are larger for those cared for by grandparents than parents (*Hypothesis 4b*).

4 After-School Tutoring Experiment

4.1 Program Description and Random Assignment

To examine the hypotheses that the model predicts for the effects of increased school tutoring, we conducted a randomized after-school tutoring intervention in Longhui County in Hunan Province of China, which was designated as a national poverty county and has a high prevalence of left-behind children. In 2010, the county's per capita GDP was only a quarter of the national average,[16] 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.[17]

We implemented the intervention as a peer-tutoring program in which high-achieving 4^{th} and 5^{th} graders were paired, respectively, with low-achieving 2^{nd} and 3^{rd} graders of the same primary school to offer one-on-one tutoring after school hours. 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) medians of their classes. Figure 2 illustrates the detailed process for the selection of tutees and tutors for the peer-tutoring program.[18] With help from the county's educational bureau, we recruited 32 primary schools to participate in the peer-tutoring program. For grades 2 and 3, 24 schools had multiple classes in each grade and eight schools had only a single class. Thus, we had a total of 64 target junior school-grades, of which 48 had multiple classes and 16 had only a single class. For each target junior school-grade with multiple classes, we randomly assigned half of the classes (a total of 60 classes) to experimental classes – from which the tutees were selected – and the remaining half (a total of 59 classes) to control classes.[19]. However, for the 16 target junior school-grades with only a single class, all of these classes were designated to be experimental classes as all schools

[16] In 2010, the county's per capita GDP was RMB 6,992, compared to the national average of RMB 29,748.

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

[18] There were two types of primary schools in this county: satellite primary schools with only junior grades (usually grades 1-3 or grades 1-4) and complete primary schools with all grades (i.e., grades 1-6). To implement this cross-age, peer-tutoring program as a randomized experiment, we had to exclude all satellite primary schools and some complete primary schools with relatively small enrollments. After applying these restrictions, we were left with a small number of schools to randomly select at the school level. The 32 participating primary schools account for about one-third of the total primary school enrollment of the county.

[19] The number of classes in each target 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.

were assured to have the opportunity to participate in the peer-tutoring program when they were recruited to join the study. Therefore, we ended up with a total of 76 experimental classes, of which 60 classes were randomly selected and had at least one control class from the same school-grade.

For each of the 76 experimental class, we randomly selected 10 low-achieving students (i.e., those with baseline scores below the class median) to participate in the peer-tutoring program as tutees, whereas the remaining unselected low-achieving students from the same classes were assigned as within-class controls. In addition, for the subset of 60 experimental classes randomly selected from the school-grades with multiple classes, all low-achieving students in the unselected control classes acted as (within-school-grade) between-class controls. In the empirical analysis of the treatment effects of the tutoring program on tutees, we employ three alternative empirical strategies: (i) a comparison between 760 randomly selected tutees and their unselected within-class controls in all the 76 experimental classes; (ii) a comparison between 600 tutees and their within-class controls from a subsample of 60 experimental classes randomly selected from 48 target junior school-grades with multiple classes; and (iii) a comparison between the 600 tutees in (ii) and their between-class controls in the unselected control classes from the same school-grades.

For each experimental class, the school assigned a senior grade class for us to recruit tutors from. We targeted only high-achievers (i.e., those with baseline scores above the class median) in these senior classes to be tutors. In order to attract as many high-achievers as possible 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 the service. Thanks to this arrangement and the cooperation of the head teachers of the senior classes, for every senior class the number of high-achieving students who applied to act as tutors under the consent of their parents/guardians exceeded the quota needed. Therefore, for each senior class we randomly selected 10 tutors from the pool of oversubscribed eligible applicants.

This tutoring experiment lasted for 8 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 a 45-minute tutorial Mondays through Thursdays.[20] 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' roles were rather passive: they only helped to keep discipline and answered questions per request, and were not supposed to be involved in any classroom teaching directly. During the tutorial sessions, the tutors helped their assigned tutees to finish homework and answered any study questions raised by the tutees.

While this experiment started with 760 assigned tutor-tutee pairs, only 90% of these pairs lasted until the end. While 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 analysis, we only used the initially assigned tutee or tutor status.

4.2 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 2013-2014 school year.

A baseline test on math and reading was conducted in September 2013, a month before

[20]The effective tutoring time is less than 45 minutes per tutorial session as sometimes the tutors and tutees worked on their own homework separately and the tutees sought help from the tutors only when they encountered questions.

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 2^{nd} and 3^{rd} graders or tutors for 4^{th} and 5^{th} 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 in every classroom.

Table 1 checks the balance of the baseline individual characteristics between tutees and controls. Parental absence is indeed a pervasive phenomenon in this sample of low achievers: Column 1 shows that of the 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 233 minutes per week, which was even more intense than the dosage of our experiment of 180 minutes per week. However, since the majority of the tutees received no tutoring help at all from home, the unconditional average weekly home-tutoring time was only 71 minutes. In all analyses of this paper, 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, these tutees on average scored 0.70σ and 0.69σ below the mean of the control classes in 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, and baseline scores. Columns 3-5 perform the balance checks comparing the subsample of 600 tutees from 60 experimental classes randomly selected from school grades with multiple classes to both their within-class and between-class controls. Except for a marginally significant difference (3.7 percentage points at 10% significance level) in the proportion being left behind by both parents between tutees in this subsample and their between-class controls,[21] other pre-experiment covariates are all

[21] In our empirical estimates, we always include specifications controlling for pre-experiment individual covariates including whether a student was left behind by both parents in the baseline.

balanced between tutees and the two control groups.

Table 2 compares the test-taking status in the endline test between the tutees and controls. Among the full sample of 760 tutees, 703 (or 92.5%) took the endline test, which is 2.4-percentage points lower (Column 1) than their within-class controls from the same experimental classes. While 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 teacher of the experimental class or the supervisor of the tutorial session had discouraged/restricted some tutees who had made the poorest progress in the program from taking the endline test, our estimated achievement effects of the program would be biased upward. The inclusion of baseline individual controls in Column 2 has little effect on the estimated difference in the observability of endline test scores, which remains significant at the 5% level. However, the coefficient on the dummy indicator for living without both parents in the baseline (-0.024) is negative and significant at the 10% level, suggesting that these left-behind children were less likely to have taken the endline test than their counterparts living with parent(s), which is not surprising since some of these left-behind children may have switched schools during our intervention either because of migrating with their parents or the changing of their guardians (e.g., from paternal grandparents to maternal grandparents). While differential selection by initial parental absence status per se is not a problem, concern may arise over the validity of comparing the treatment effects between subgroups of tutees defined by initial parental absence status if being selected as a tutee in our after-school tutoring program had affected the chances of taking the endline test for children left behind by both parents differently than children living with parent(s). Thus, in Column 3 we further include an interaction term between the treatment status and a dummy indicator for living without both parents in the baseline. The coefficient on the interaction term is quantitatively very small (-0.011) and statistically insignificant, suggesting no evidence to substantiate such a concern. When a subsample of 600 tutees from schools with multiple classes in each junior grade is compared to their within-class and between-class controls in Columns 4-6 and 7-9, the differences in the test taking status becomes smaller (though still negative) and statistically insignificant. In the next section examining the achievement effects of the after-school tutoring experiment, we first conduct the main analysis without considering the potential sample selection problem. Then, for the estimates using the full sample of 76 experimental classes, for which the difference in sample selection between tutees and controls is indeed significant, we perform a partial identification analysis to construct bounds on the estimated treatment effects (Section 5.4.2).

5 Results on Student Achievement

5.1 Average Achievement Effects for Tutees

Our main empirical strategy to assess the average achievement effect of the after-school tutoring program on the tutees uses the following class fixed-effect regression applied to both tutees and their within-class controls in the experimental classes:

$$\Delta y_{ij} = y_{ij1} - y_{ij0} = \lambda D_{ij} + X_{ij}\beta + \varphi_j + \varepsilon_{ij}, \tag{6}$$

where Δy_{ij} denotes the change in the test scores of student i from class j between the baseline and endline tests, D_{ij} is a dummy indicator equal to 1 if student i was assigned to be a tutee and 0 if otherwise, X_{ij} is a vector of control variables including baseline test scores, child gender, and parental absence status in the baseline, φ_j is a class fixed effect that captures the unobserved determinates of learning shared in common at the regular classroom among all students from class j, and ε_{ij} is an error term, consisting of both an individual-level component and a class-level component. This specification assumes that there are no spillover effects of the tutoring on the controls. But spillover effects, probably positive, are a possibility since any improved performance by the tutees due to the program might benefit their classmates. If there were positive spillover effects, the estimate of the treatment effect would be a lower bound, with no obvious reason for differential biases by parental status (considered in Section 5.2).

Alternatively, we can also assess the effect of the after-school tutoring program on the tutees with a school-grade fixed-effect regression comparing tutees with their between-class controls (i.e., other low-achieving students from control classes in the same school grades) as follows:

$$\Delta y_{ijg} = \lambda D_{ijg} + X_{ijg}\beta + \phi_g + \nu_{ijg}, \tag{7}$$

where Δy_{ijg} denotes the change in the test scores of student i from class j of school-grade g between the baseline and endline tests, ϕ_g is a school-grade fixed effect that captures the unobserved determinants of learning shared in common in the school-grade among all students from school-grade g, and the error term ν_{ijg} consists of both an individual-level component and a class-level component. Note that because both ε_{ij} and ν_{ijg} have class-level components, we always cluster the standard errors at the class level in estimates where the dependent variable involves test scores.

Panel A of Table 3 reports estimates of the average treatment effect on tutees' math scores. Columns 1-3 estimate the class fixed-effect regressions employing the full sample of tutees and their within-class control students from 76 experimental classes. Column 1 includes no control variables, Column 2 controls for baseline scores only, and Column 3 further controls for baseline individual characteristics including gender and parental absence status. The point estimates of the coefficient λ change relatively little with the inclusion of additional control variables. Taking the specification in Column 3 with the full set of control variables as an example, the estimated coefficient indicates that the tutees in the full sample of 76 experimental classes have an average score gain of 0.136σ in math (significant at 1% level) compared to the control students from the same classes. Columns 4-6 replicate the same class fixed-effect estimates in Columns 1-3 but use only a subset of 60 experimental classes randomly selected from the school grades with multiple classes. The estimates for this subsample are somewhat smaller compared to the full sample, but still significant at the 5% level. Column 7-9 estimate the school-grade fixed-effect regressions in Equation (7) employing the tutees from the subset of the 60 randomly selected experimental classes and the control students from the unselected control classes within the same school grades. These school-grade fixed-effect estimates (0.087-0.110σ) are quantitatively very similar to those class fixed-effect estimates (0.090-0.102σ) obtained in Columns 4-6 for the same subsample of tutees, but are less precisely estimated.[22] Nonetheless, when baseline scores are controlled

[22]The reduction in the statistical precision of the school-grade fixed-effect estimates compared to the class fixed-effect estimates is largely caused by the clustering of the standard errors at the class level. This is

for in Columns 11-12, these school-grade fixed-effect estimates remain significant at the 10% level.

Panel B of Table 3 reports estimates of the average treatment effect on tutees' reading scores. Regardless of the empirical specifications and the samples 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. One may well expect that the value of outside tutoring may differ across subjects. For example, in mathematics, there are many problem-solving "tricks", which can be best learnt through individual tutoring. For language, on the other hand, a student can comprehend new words only if he/she spends effort to memorize them. Thus, if individual tutoring increases test scores largely through improving learning and test-taking skills, which are more important for math than reading, the achievement effect of individual tutoring would also be more salient for math than reading. 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., Abdulkadiroğlu et al., 2011; Angrist et al., 2010, 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 U.S. with the first 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.

For our tutoring intervention, moreover, the 4^{th} and 5^{th} graders are likely to be much less effective tutors for reading than math. In that case, both A and θ in the achievement production function Equation (1) could be subject-specific and much smaller for reading than for math. Moreover, anecdotal evidence collected from the teachers supervising the tutorial rooms also indicates that in general the pairs spent more time working on math than reading during the tutorial sessions.[23] Because we never obtain any statistically significant

because when tutees and their controls from the same experimental class are used to estimate Equation (6), the variation in the treatment status is at the individual level, so clustering at the class level has relatively little effect on the standard error of the estimated coefficient. However, when tutees from the experimental classes and other low-achieving students from the control classes are used to estimate Equation (7), the treatment status only varies at the class level. As a result, clustering at the class level significantly increases the standard error of the estimated coefficient.

[23]Differential tutoring time spent on math and reading may be due to the nature of the homework for

or quantitatively large estimates for reading scores (results available upon requests), for the more detailed analysis in the remaining of the paper we focus only on the effects of the program on tutees' math scores.

5.2 Differential Achievement Effects for Tutees by Parental Absence Status

The positive and significant estimates of the coefficient λ in Panel A of Table 3 are consistent with *Hypothesis 3* that the after-school tutoring program resulted in net achievement gains as far as math scores are concerned. *Hypothesis 4b* further predicts achievement effects to be greater for students from households facing higher unit costs of home tutoring. In this subsection, we use children's initial parental absence status to proxy for households' own costs of home tutoring and investigate the heterogeneity in the math achievement effects for tutees with varying parental absence status in the baseline to test *Hypothesis 4b*.

To allow the treatment effects to vary by tutees' parental absence status, we first interact the treatment dummy (D_{ij}) with three mutually exclusive dummy indicators representing a student's initial 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 4 reports estimates of an expanded version of the class fixed-effect model in Equation (6) 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.

these two subjects. Usually, math homework requires a lot of written work and calculation whilst language homework may be more on reading and reciting that do not need to be done in front of a tutor.

Columns 3-4 report the results of the class fixed-effect estimations using the full sample of tutees employing only two interaction terms $D_{ij} \times NoParent_{ij}$ and $D_{ij} \times Parent_{ij}$. The estimates in Column 3 suggest that the intervention 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. While 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 into a combined category. Moreover, consistent with *Hypothesis 4b*, the difference in these point estimates (0.120σ, with a p-value of 0.02) indicates that the after-school tutoring program yielded a larger test score gain for tutees cared for by grandparents than those cared directly by parent(s).[24]

Columns 5-6 and 7-8 of Table 4 report, respectively, the class and school-grade fixed-effect estimates for the subsample of tutees from 60 randomly selected experimental classes. Though always positive, the coefficients are less precisely estimated for this subsample. For tutees living with parent(s), the reduction in the precision of these estimates make us unable to reject the null hypothesis that the program does not generate significant achievement gains for them, although we also cannot reject that the achievement gains are the same in magnitude as the statistically significant estimates obtained for the full sample in Columns 3-4 (0.083-0.094σ). Nonetheless, for tutees living without both parents, the point estimates (0.122-0.149σ) remain statistically significant in all estimates, although somewhat smaller in magnitude than those obtained for the full sample. Taken together, the results in Table 4 point to larger math achievement gains for tutees living without both parents, suggesting an important role of family background (as measured by parental absence status here) in affecting the efficacy of the after-school tutoring intervention implemented. In Section 6, we further explore the mechanisms that result in differential math score gains for tutees with varying parental absence status by comparing changes in their home-tutoring inputs after the treatment.

[24] If we compare the coefficients in Column 4, the difference is quantitatively somewhat smaller (0.096σ) and only marginally significant at the 15% level. However, the magnitudes of these point estimates (0.190σ vs. 0.094σ) still indicate that the achievement effect for tutees living without both parents is twice as large as that for those living with parent(s).

5.3 Robustness Analysis

5.3.1 Accounting for Other Dimensions of Potential Heterogeneity in Treatment Effects

In an earlier study, we document significant adverse effects of the absence of both parents on children's contemporary cognitive development (Zhang et al., 2014). To the extent that a child's parental absence status observed in the baseline is correlated with his/her parents' past migration histories, children observed to be left behind by both parents in the baseline would on average have a large disadvantage in the cumulative parenting inputs received during their growing-up years compared to those living with parent(s) in the baseline. Since we selected only students who had scored below their class medians in the baseline test to be eligible for the after-school tutoring program, it is possible that conditional on having scored below the class median in the baseline test those left behind by both parents on average had higher innate ability than those living with parent(s). That is, some children left behind by both parents were low-achievers because of their disadvantages in family educational inputs rather than in innate cognitive abilities. Thus, a potential reason for the tutoring program to yield larger math score gains for children left behind by both parents is the complementarity between the tutoring inputs received from our intervention and children's innate cognitive ability. If that is indeed the mechanism for differential treatment effects by parental absence status, one would also expect the after-school tutoring program to yield larger test score gains for the relatively higher-achieving tutees in the baseline test conditional on their parental absence status in the baseline.

Column 1 of Table 5 tests this hypothesis by adding an interaction term between the treatment dummy (D_{ij}) and student's baseline test score (Y_{ij0}) to the class fixed-effect specification in Equation (6). The coefficient on this interaction is negative and insignificant, showing no support for larger achievement gains for tutees with higher baseline scores, which proxies for higher innate cognitive ability. The coefficient on the treatment dummy itself (0.122σ) is statistically significant and quantitatively similar to the estimate (0.136σ) obtained without including the interaction term in Column 3, Panel A of Table 3. In Column 2 of Table 5, we retain the interaction term between the treatment status and baseline score ($D_{ij} \times Y_{ij0}$), but replace the treatment dummy (D_{ij}) by its interactions with the two dummy indicators for living with one or both parents ($Parent_{ij}$) and without both parents ($NoParent_{ij}$). The coefficient on the interaction term $D_{ij} \times Y_{ij0}$ remains small, negative, and insignificant, whereas those on the interaction terms $D_{ij} \times NoParent_{ij}$ (0.177σ) and $D_{ij} \times Parent_{ij}$ (0.084σ) are similar in size to the corresponding estimates (0.190σ and 0.094σ, respectively) obtained in Column 4 of Table 4 without the inclusion of the interaction term $D_{ij} \times Y_{ij0}$. However, a rather small reduction in the point estimate (from 0.094σ to 0.084σ) makes the coefficient on $D_{ij} \times Parent_{ij}$ no longer significant at conventional levels, although the coefficient on $D_{ij} \times NoParent_{ij}$ (0.177σ) remains significant at the 1% level.

In Columns 3-4 of Table 5, we investigate the heterogeneity in the treatment effects on math scores by gender. We find no evidence that the treatment effects differ between boys and girls: the coefficient estimates on the interaction terms between the treatment dummy and the female dummy are always small and insignificant. The inclusion of these interaction terms also has little effect on the estimated overall average treatment effect (0.123σ) in Column 3 and the respective average treatment effect for tutees living with at least one parent (0.084σ) and without both parents (0.178σ) in Column 4. In Columns 5-6 of Table 5, we further include the separate interactions of the treatment dummy with baseline scores and gender. None of the coefficients on these interaction terms is significant or large in magnitude. Although the precision of the estimates is somewhat reduced, the qualitative conclusions remain robust to the inclusion of these additional interaction terms to account for differences in the treatment effects by baseline scores and gender. In Column 5, the estimated coefficient on the treatment dummy itself (0.110σ) remains significant at the 10% level and is similar in magnitude to those obtained earlier without controlling for these interaction terms. In Column 6, the estimated achievement effect for tutees living with no parent (0.167σ) remains significant at the 5% level and is twice as large as that for those living with parent(s) (0.075σ).

5.3.2 Partial Identification Analysis

So far our analysis of the achievement effects of the program has ignored the potential sample selection in taking the endline test. However, the results in Table 2 seem to indicate that the control students might be somewhat more likely to have taken the endline test than the tutees. Although the differences in their test-taking rates are small in magnitude (no more than 2.5 percentage points), it is statistically significant at the 5% level among the full sample of 760 tutees and 1,184 control students from the 76 experimental classes (Columns 1 2, Table 2). To check the robustness of our results to this potential differential sample selection problem, we implement a trimming procedure proposed by Lee (2009), which under the assumption of a monotone effect of the treatment on sample selection would yield conservative bounds on the treatment effects on tutees' math scores in the presence of differential sample selection. [25] Since the difference in sample selection between tutees and controls is only significant for the full sample of tutees and their within-class controls, we apply this trimming procedure to this sample only.

Among the full sample of 760 assigned tutees and 1,184 assigned controls from the 76 experimental classes, we observe the endline test scores for 703 tutees and 1,112 controls. However, had the test-taking rates been the same between the tutees and controls from the same class, we would expect 30 fewer control students $(1,184 \times 0.025)$ taking the endline test. Thus, the lower- and upper-bounds on the treatment effects can be constructed by trimming the bottom and top tails of the distribution of the outcome variable (i.e., changes in math scores between the baseline and endline tests) by this number. Our strategy to identify the set of control students pertaining to the bottom and top tails to be trimmed is as follows. First, from a total of 76 experimental classes, we randomly select 30 classes and label them as trim-target classes. Second, for each trim-target class, we rank all control students in descending order of changes in math scores between the baseline and endline tests (Δy_{ij}). Third, we trim the bottom-ranked (or top-ranked) control student from each of the 30 trim-target classes and obtain the lower-tail-trimmed (or upper-tail-trimmed) data for constructing the lower-bound (or upper-bound) estimates.

Columns 1-3 in Panel A of Table 6 report the lower-bound estimates using the lower-trimmed data corresponding to the same estimates in Columns 1-3 in Panel A of Table 3 using the untrimmed data. Although by construction these lower-bound estimates (0.101-0.110σ)

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

are smaller in magnitude, they remain statistically significant at the 1% level, showing that our finding of a positive and significant overall average treatment effect on tutees' math scores is unaffected by the small degree of differential sample selection between tutees and controls taking the endline test. Columns 4-6 use the lower-trimmed data to estimate differential treatment effects by whether a student was living with any parent or not in the baseline. Not surprisingly, for both types of tutees, the estimated treatment effects using these lower-trimmed data are smaller than those obtained from the untrimmed data in Table 4. Because of the reduction in size, the estimates for tutees living with parent(s) are no longer statistically significant. However, the estimates for tutees living without both parents remain significant at the 1% level and more than twice as large as those for tutees living with parent(s) in all specifications. Panel B of Table 6 reports the corresponding upper-bound estimates employing the upper-trimmed data. All estimates are larger in size compared to the corresponding estimates using the untrimmed data. When differential treatment effects by parental absence status are examined in Columns 4-6, the estimates are significant for both types of tutees, but the estimates for tutees living without both parents remain at least twice as large as for those living with parent(s).

5.4 Achievement Effects for the Tutors

As Figure 2 illustrates, the tutors were randomly selected from a pool of 1,405 eligible applicants from 76 senior classes who had scored above their class medians in the baseline test and also obtained their guardians' consent to act as tutors in the after-school tutoring program. Therefore, for each senior class that the tutors were recruited from, the eligible applicants who were not selected to be tutors in our program can be used as controls. In Table 7, we estimate the treatment effects of our program on both the math (Columns 1-3) and reading scores (Columns 4-6) for the tutors using their unselected eligible classmates as the control group. For both subjects, the estimated average treatment effects are statistically and quantitatively indistinguishable from 0 (Columns 1 and 4), and are unaffected by the inclusion of individual control variables (Columns 2 and 5). However, the coefficients on the female dummy are significant for both subjects but are opposite in sign (negative for math and positive for reading), suggesting the divergence in the patterns of the achievement dynamics between girls and boys in the two subjects during senior primary school years. In Columns 3 and 6, we further include the interactions of the tutor dummy with inclusion of additional interaction terms between the tutor dummy with gender and baseline scores, none of the coefficients on the interaction terms is significant, showing no evidence for any heterogeneous treatment effects for the tutors by gender or initial achievement. Any negative effects on tutors' test scores from the greater demands on their time due to tutoring, thus, seem to be offset by positive gains from reviewing and explaining material to tutees.

6 Results on Private Household Inputs

Recall that the model in Section 3 predicts that parents invest more in home-tutoring inputs than grandparents (*Hypothesis 1*) but also reduce their home-tutoring inputs more when the amount of tutoring received in school increases (*Hypothesis 4a*). We test these hypotheses empirically in this section. We first examine the difference in the baseline levels of home-tutoring inputs by children's varying parental absence status before the intervention (Section 6.1). To understand the mechanisms behind the substantially larger math score gains for tutees living without both parents than those living with parent(s), we further compare changes in home-tutoring inputs by students' treatment status in the experiment and parental absence status (Section 6.2).

6.1 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 relationship between parental absence status and baseline home-tutoring inputs:

$$H_{ij0} = \alpha_0 NoParent_{ij} + \alpha_1 OneParent_{ij} + \alpha_2 TwoParents_{ij} + \mu_j + \eta_{ij}, \tag{8}$$

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 time 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 Section 5.2, and μ_j is a class fixed effect. In estimating Equation (8), 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 α_0, α_1, and α_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}.

Column 1 of Table 8 reports estimates of Equation (8) when H_{ij0} is measured by the dummy indicator for having received any tutoring at home in the week prior to the baseline survey. For children living with one and both parents, respectively, the proportions who had received tutoring at home in the baseline are almost exactly the same (35.6% and 35.7%), both of which are substantially higher than that of children living without both parents (27.1%). Column 3 of Table 8 reports estimates of Equation (8) when H_{ij0} is measured by the total weekly home-tutoring time. The baseline home-tutoring time is also very similar between children living with a single parent (89.6 minutes/week) and both parents (92.2 minutes/week), but is much smaller for those living without both parents (61.1 minutes/week).[26]

Since the absence of a single parent has little effect on the home-tutoring inputs that children received in the baseline, we pool children living with one and both parents together and estimate the following model:

$$H_{ij0} = \alpha_0 NoParent_{ij} + \alpha_1 Parent_{ij} + \mu_j + \eta_{ij}, \tag{9}$$

where $Parent_{ij}$ is a dummy equal to 1 if student i was living with one or both parents as defined before. Under the assumption that parents have comparative advantage in home tutoring over grandparents, *Hypothesis 1* in Section 3 is that parents choose higher levels of home-tutoring inputs than grandparents, i.e., $\alpha_1 > \alpha_0$. We test this hypothesis empirically using our estimates of the coefficients for Equation (9). The testing results show that children

[26]It is worth noting that the differences in the observed home-tutoring time between children living with at least one and without both parents may underscore the differences in the effective tutoring inputs that they received at home if one further takes into account the differences in the efficacy of home tutoring offered by parents and grandparents.

living with parent(s) were 8.5 percentage points (or 31.3%) more likely to report tutoring at home (Column 2, Table 8) and on average reported 29.7 minutes' (or 48.6%) more home-tutoring time than those left-behind by both parents (Column 4, Table 8). Both the extensive and intensive margin differences are significant at the 1% level.

6.2 Changes in Home-Tutoring Inputs

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}), i.e., $H_{ijt} = H^*_{ijt} + e_{ijt}$ for $t \in \{0, 1\}$. For H_{ij0} in the baseline, no student was aware of the after-school tutoring program to be implemented and thus should have had no strategic incentive to manipulate their reported home-tutoring time. Any reporting error (if it exists) should be balanced between the tutees and controls because of the random assignment, i.e., $E\left[e_{ij0}|D_{ij} = 1\right] = E\left[e_{ij0}|D_{ij} = 0\right]$. However, for H_{ij1} in the endline, the implementation of the after-school tutoring program may have altered the incentives and reporting behaviors of both the tutees and controls, but to different degrees.[27] As a result, there may be some systematic difference in the reporting error between the two groups, i.e., $E[e_{ij1}|D_{ij} = 1] \neq E[e_{ij1}|D_{ij} = 0]$. To the extent that the reporting behaviors of the tutees and controls were affected differently in the endline, the class fixed-effect specification used to estimate the program effects on test scores in Equation (6) cannot be applied directly to estimate the program effects on home-tutoring inputs.

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:

$$\Delta H_{ij} = H_{ij1} - H_{ij0} = \pi_0^\kappa NoParent_{ij} + \pi_1^\kappa Parent_{ij} + \theta_j^\kappa + \xi_{ij}^\kappa, \kappa \in \{T, C\}. \qquad (10)$$

The superscript $\kappa \in$ {T,C} denotes whether the equation is estimated using the tutees

[27]Although the schools had explained that access to the after-school tutoring program was determined by a lottery, some unselected families (i.e., those in the control group) still complained for not having the opportunity to participate in the after-school tutoring program. Some families may think that the program targeted students with relatively fewer home-tutoring inputs and thus may have incentives to underreport their home tutoring time in the endline.

subsample (T) or the controls subsample (C). Note that the class fixed-effect δ_j^κ is now normalized to sum to 0 across all students in each subsample. Column 1 of Table 9 reports estimates of Equation (10) for the subsample of tutees. For tutees living with parent(s), the estimates of the coefficient on $Parent_{ij}$ dummy (π_1^T) indicate that they were 16.4 percentage points less likely to report having received tutoring at home in the endline (Panel A) and on average reported 30.3 minutes' less home-tutoring time per week compared to the baseline (Panel B). However, for tutees living apart from both parents, the estimates of the coefficient on $NoParent_{ij}$ dummy (π_0^T) indicate much smaller and statistically insignificant reductions in the self-reported home-tutoring time at both the extensive and intensive margins. Because of the existence of an ambiguous effect of participating in the after-school tutoring program on reporting error e_{ij1} for the tutees, we cannot interpret these estimates as reflecting changes in the actual home-tutoring inputs (ΔH_{ij}^*). However, as long as the effect of program participation on the reporting behavior is the same for tutees with different initial parental absence status, i.e., $E\left[e_{ij1} \middle| Parent_{ij}=1, D_{ij}=1, \delta_j^T\right] = E\left[e_{ij1} \middle| NoParent_{ij}=1, D_{ij}=1, \delta_j^T\right]$, then the difference in the two coefficients $\pi_1^T - \pi_0^T$ still reflects the differential response in the *actual* home-tutoring inputs (ΔH_{ijt}^*) between tutees living with and without parent(s), i.e., $\pi_1^T - \pi_0^T = E\left[\Delta H_{ij}^* \middle| Parent_{ij}=1, D_{ij}=1, \delta_j^T\right] - E\left[\Delta H_{ij}^* \middle| NoParent_{ij}=1,\ D_{ij}=1, \delta_j^T\right]$. Recall that *Hypothesis 4a* in Section 3 predicts that parents reduce their home-tutoring inputs more than grandparents, i.e., $\pi_1^T - \pi_0^T < 0$. We test this hypothesis formally at the bottom of each panel in Column 1 of Table 9. For both measures of home-tutoring inputs, the difference $\hat{\pi}_1^T - \hat{\pi}_0^T$ 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 parent(s).

Column 2 of Table 9 reports estimates of Equation (10) 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 self-reported home-tutoring inputs is likely to be dominated by the effect on their reporting behavior in the endline (e_{ij1}) rather than actual home-tutoring inputs (H_{ij1}^*). Moreover, if the program *nonparticipation* effect on reporting behavior is the same by initial parental absence status, i.e., $E\left[e_{ij1} \middle| Parent_{ij}=1, D_{ij}=0, \delta_j^C\right] =$

$E\left[e_{ij1} \middle| NoParent_{ij} = 1, D_{ij} = 0, \delta_j^C\right]$, then the overall effect on changes in the self-reported home-tutoring inputs (ΔH_{ij}) should also not differ significantly between control students living with parent(s) and those living without both parents. Indeed, both groups of control students reported large and significant reductions in home-tutoring time, and the difference in the extent of the reduction is rather small quantitatively: -17.0 vs. -14.4 percentage points at the extensive margin and -40.4 vs. -33.6 minutes at the intensive margin. Moreover, we cannot reject the possibility that changes in the self-reporting home-tutoring inputs are the same between the two groups of control students (i.e., $\pi_1^C = \pi_0^C$), suggesting no evidence that the existence of 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 Column 1 of Table 9 indeed identifies the differential response in the actual home-tutoring inputs for tutees with different parental absence status in the baseline.

Note that because we estimate Equation (10) separately for the tutees and controls, the class fixed effects δ_j^κ generally differ between the two estimates (i.e., $\delta_j^T \neq \delta_j^C$). To check the sensitivity of the results to the way that the class fixed effects are accounted for, we also estimate an alternative specification pooling together the treated and control students as follows:

$$
\begin{aligned}
\Delta H_{ij} = {} & \pi_0^T (D_{ij} \times NoParent_{ij}) + \pi_1^T (D_{ij} \times Parent_{ij}) \\
& + \pi_0^C \Big((1 - D_{ij}) \times NoParent_{ij} \Big) + \pi_1^C \Big((1 - D_{ij}) \times Parent_{ij} \Big) + \delta_j + \xi_{ij},
\end{aligned}
\tag{11}
$$

where the class fixed effect δ_j is normalized to sum to 0 across all students including both tutees and controls. Columns 3-4 of Table 9 report estimates of this pooled regression. For the control students, the results are qualitatively the same as those found in the separate estimation in Column 2: regardless of the measure of ΔH_{ij} used, both $\hat{\pi}_1^C$ and $\hat{\pi}_0^C$ are negative and significant and their difference $\hat{\pi}_1^C - \hat{\pi}_0^C$ is always small in magnitude and statistically insignificant (Panels A and B, Column 4). For the tutees, the results also remain similar to those found in the separate estimates in Column 1, although for some coefficients the statistical significance is changed due to quantitative differences in the point estimates. When ΔH_{ij} is measured by the change in the dummy indicator for self-reported home tutoring (Panel A, Column 3), the estimate of the reduction for tutees living without parents ($\hat{\pi}_0^T = -0.060$) becomes statistically significant itself, although its difference with the reduction for tutees living with parent(s) remains statistically significant ($\hat{\pi}_1^T - \hat{\pi}_0^T = -0.092$). When ΔH_{ij} is measured by the change in self-reported weekly home-tutoring time (Panel B, Column 3), the estimate for the difference $\pi_1^T - \pi_0^T$ is somewhat reduced in magnitude to -20.0 minutes (with a p-value of 0.137), and is no longer significant at conventional levels. Nonetheless, the results of the varying exercises reported in Table 9, taken together, are by and large consistent with *Hypothesis 4a* that parents reduce their home-tutoring inputs more than grandparents in response to the increase in school tutoring. The smaller extent of the reduction in home-tutoring inputs for tutees left-behind by both parents is also likely to be the main reason for them to experience larger math score gains from participating in the after-school tutoring program than those living with parent(s).

7 Conclusion

Worldwide students' access to learning enrichment activities after school is highly dependent on family backgrounds, 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 academically lagging behind compared to their counterparts 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 home learning support after school (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 investigating possible compensatory programs for ameliorating 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 education inequality, prior evidence of the effects of public after-school programs on children's academic outcomes is limited and far from unified.

This paper develops a simple model of student achievement that integrates public and private after-school tutoring inputs to determine student achievement and presents empirical evidence consistent with the model using a randomized after-school tutoring experiment in a poor area of rural China in which high-achieving 4^{th} and 5^{th} graders provided high-dosage, one-on-one tutoring to low-achieving 2^{nd} and 3^{rd} 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.

Thus our results demonstrate that peer tutoring is a feasible and 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 backgrounds on child development and educational inequality. Moreover, an important policy implication of our results that targeting public provision of after-school learning support to children from disadvantaged family backgrounds, such as children living with no parents in rural China, is both an equitable and efficacious strategy as these children also tend to lag behind in academic achievement and benefit the most from such public interventions. Last but not least, although we only implemented a particular form of after-school learning support (i.e., 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 e-learning resources that more easily can be scaled up.

References

Abdulkadiroğlu, Atila, Joshua D Angrist, Susan M Dynarski, Thomas J Kane, and Parag A Pathak. 2011. "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots." *Quarterly Journal of Economics*, 126(2): 699–748.

Aizer, Anna, and Flavio Cunha. 2012. "The Production of Human Capital: Endowments, Investments and Fertility." *NBER working paper, NO.18429.*

All-China Women's Federation. 2014. *A Study of the Situation of Left-behind Children in Rural China (in Chinese).*

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

Angrist, Joshua D, Susan M Dynarski, Thomas J Kane, Parag A Pathak, and Christopher R Walters. 2010. "Inputs and Impacts in Charter Schools: KIPP Lynn." *American Economic Review*, 100(2): 239–43.

Angrist, Joshua D, Susan M Dynarski, Thomas J Kane, Parag A Pathak, and Christopher R Walters. 2012. "Who Benefits from KIPP?" *Journal of policy Analysis and Management*, 31(4): 837–860.

Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy*, 2(1): 1–30.

Banerjee, Abhijit V, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics*, 122(3): 1235–1264.

Bitler, Marianne P, Hilary W Hoynes, and Thurston Domina. 2014. "Experimental Evidence on Distributional Effects of Head Start." *NBER woking paper, NO.20434.*

Black, Alison Rebeck, Fred Doolittle, Pei Zhu, Rebecca Unterman, and Jean Baldwin Grossman. 2008. "The Evaluation of Enhanced Academic Instruction in After-School Programs: Findings After the First Year of Implementation." NCEE 2008-4021, U.S. Department of Education.

Bray, Mark. 2007. *The Shadow Education System: Private Tutoring and Its Implications for Planners (Second Edition).* Paris:UNESCO International Institute for Educational Planning.

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

Cabezas, Verónica, José I Cuesta, and Francisco A Gallego. 2011. "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.

Cascio, Elizabeth U, and Diane Whitmore Schanzenbach. 2013. "The Impacts of Expending Access to High Quality Preschool Education." *Brookings Papers on Economic Activity*, 161–193.

Chandler, Michael Alison. 2011. "S. Korea Tries to Wrest Control from Booming Private Tutoring industry." *Washington Post (April 3, 2011).* https://www.washingtonpost.com/world/s-korea-tries-to-wrest-control-from-booming-private-tutoring-industry/2011/01/12/AFNXQfXC_story.html?noredirect=on(accessed 18 April 2019).

Cook, Philip J, Kenneth Dodge, George Farkas, Roland G Fryer, Jonathan Guryan, Jens Ludwig, Susan Mayer, Harold Pollack, and Laurence Steinberg. 2014. "The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago." *NBER Working Paper, NO.19862.*

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy*, 126(6): 2356–2409.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97(2): 31–47.

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics*, 5(2): 29–57.

Dobbie, Will, and Roland G Fryer. 2013. "Getting Beneath the Veil of Effective Schools: Evidence from New York City." *American Economic Journal: Applied Economics*, 5(4): 28–60.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*, 101(5): 1739–74.

Duncan, Greg J, and Richard J Murnane. 2016. "Rising Inequality in Family Incomes and Children's Educational Outcomes." *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(2): 142–158.

Fryer, Roland G. 2014. "Injecting charter school best practices into traditional public schools: Evidence from field experiments." *Quarterly Journal of Economics*, 129(3): 1355–1407.

Gaastra, Sieuwerd. 2016. "The Effect of After-school Programs on Short-run Student Academic and Behavioral Outcomes." Unpublished manuscript, University of California, San Diego.

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

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives*, 22(3): 23–46.

Havnes, Tarjei, and Magne Mogstad. 2015. "Is Universal Child Care Leveling the Playing Field?" *Journal of Public Economics*, 127: 100–114.

Heckman, James J. 2007. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences*, 104(33): 13250–13255.

James-Burdumy, Susanne, Mark Dynarski, and John Deke. 2007. "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(4): 296–318.

Kim, Sunwoong, and Ju-Ho Lee. 2010. "Private Tutoring and Demand for Education in South Korea." *Economic Development and Cultural Change*, 58(2): 259–296.

Kottelenberg, Michael J, and Steven F Lehrer. 2017. "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Child Care." *Journal of Labor Economics*, 35(3): 609–653.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies*, 76(3): 1071–1102.

Li, Hongbin, Prashant Loyalka, Scott Rozelle, Binzhen Wu, and Jieyu Xie. 2015. "Unequal Access to College in China: How Far Have Poor, Rural Students Been Left Behind?" *The China Quarterly*, 221: 185–207.

Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle. 2014. "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools." *Journal of Public Economics*, 111: 29–45.

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

Matsuoka, Ryoji. 2015. "School Socioeconomic Composition Effect on Shadow Education Participation: Evidence from Japan." *British Journal of Sociology of Education*, 36(2): 270–290.

Meyer, Erik, and Chris Van Klaveren. 2013. "The Effectiveness of Extended Day Programs: Evidence from a Randomized Field Experiment in the Netherlands." *Economics of Education Review*, 36: 1–11.

Miller, Sarah, and Paul Connolly. 2013. "A Randomized Controlled Trial Evaluation of Time to Read, A volunteer Tutoring Program for 8- to 9-Year-Olds." *Educational Evaluation and Policy Analysis*, 35(1): 23–37.

Mills, Jonathan N, and Patrick J Wolf. 2017. "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement After 2 Years." *Educational Evaluation and Policy Analysis*, 39(3): 464–484.

Park, Hyunjoon, Claudia Buchmann, Jaesung Choi, and Joseph J Merry. 2016. "Learning Beyond the School Walls: Trends and Implications." *Annual Review of Sociology*, 42: 231–252.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and BehavioralResponses." *American Economic Review*, 103(4): 1289–1324.

Ramey, Garey, and Valerie A Ramey. 2010. "The Rug Rat Race." *Brookings Papers on Economic Activity*, 2010(1): 129–176.

Safarzyńska, Karolina. 2013. "Socio-economic Determinants of Demand for Private Tutoring." *European Sociological Review*, 29(2): 139–154.

State Council of China. 2018. "“Opinions on Regulating the Development of Off-Campus Training Institutions." *General Office of the State Council Document No. 80 (in Chinese).*

Todd, Petra E, and Kenneth I Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal*, 113(485): F3–F33.

U.S.Department of Education. 2015. *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.

Weiss, Heather B, Priscilla MD Little, Suzanne M Bouffard, Sarah N Deschenes, and Helen Janc Malone. 2009. *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.

Weiss, Michael J, Alyssa Ratledge, Colleen Sommo, and Himani Gupta. 2019. "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(3): 253–97.

Wei, Yi. 2017. *The 2017 Chinese Family Survey of Educational Expenditure (in Chinese).* China Institute of Education Finance Research. `http://ciefr.pku.edu.cn/cbw/kyjb/2018/03/kyjb_5257.shtml`(accessed 18 April 2019).

Yuan, Cheng, and Lei Zhang. 2015. "Public Education Spending and Private Substitution in Urban China." *Journal of Development Economics*, 115: 124–139.

Zhang, Dandan, Xin Li, and Jinjun Xue. 2015. "Education Inequality between Rural and Urban Areas of the People's Republic of China, Migrants' Children Education, and Some Implications." *Asian Development Review*, 32(1): 196–224.

Zhang, Hongliang. 2016. "Identification of Treatment Effects under Imperfect Matching with an Application to Chinese Elite Schools." *Journal of Public Economics*, 142: 56–82.

Zhang, Hongliang, Jere R Behrman, C Simon Fan, Xiangdong Wei, and Junsen Zhang. 2014. "Does parental absence reduce cognitive achievements? Evidence from rural China." *Journal of Development Economics*, 111: 181–195.

Zhang, Wei. 2014. "The Demand for Shadow Education in China: Mainstream Teachers and Power Relations." *Asia Pacific Journal of Education*, 34(4): 436–454.

Zhang, Yueyun, and Yu Xie. 2016. "Family Background, Private Tutoring, and Children's Educational Performance in Contemporary China." *Chinese sociological review*, 48(1): 64–82.

Zhao, Guochang, Jingjing Ye, Zhengyang Li, and Sen Xue. 2017. "How and Why do Chinese Urban Students Outperform Their Rural Counterparts?" *China Economic Review*, 45: 103–123.

Zimmer, Ron, Laura Hamilton, and Rachel Christina. 2010. "After-school Tutoring in the Context of No Child Left Behind: Effectiveness of Two Programs in the Pittsburgh Public Schools." *Economics of education Review*, 29(1): 18–28.

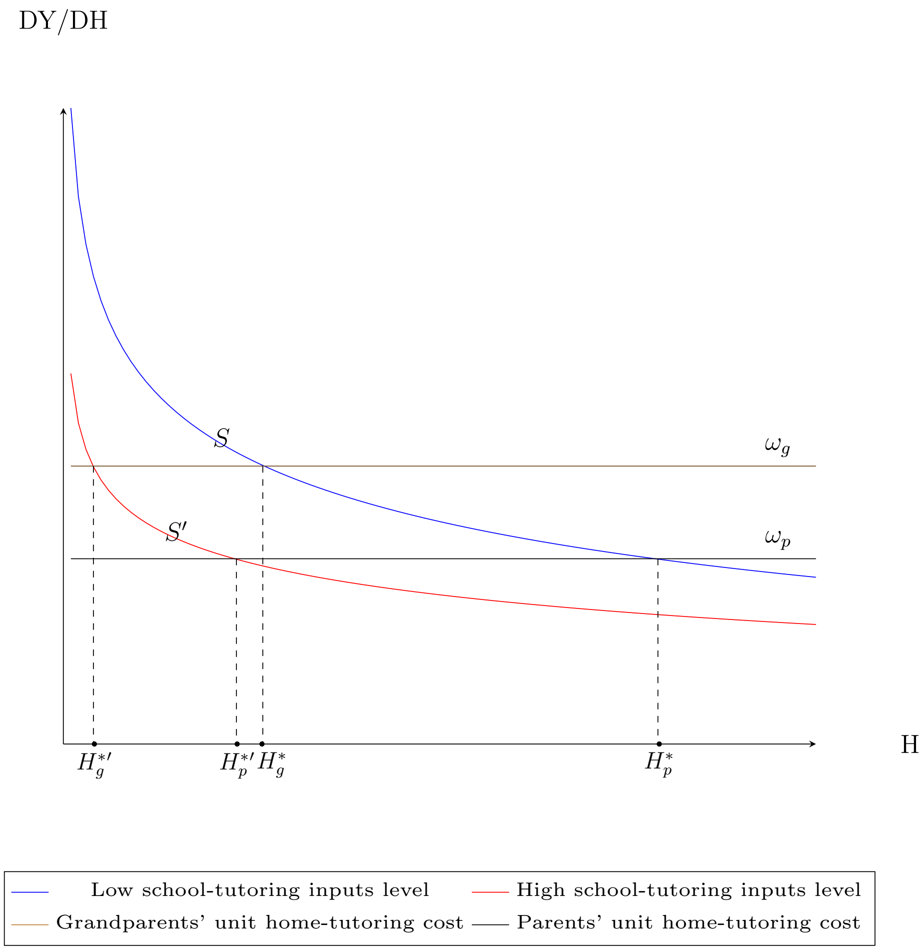

Figure 1: Household's Optimal Choices of Home-tutoring Inputs

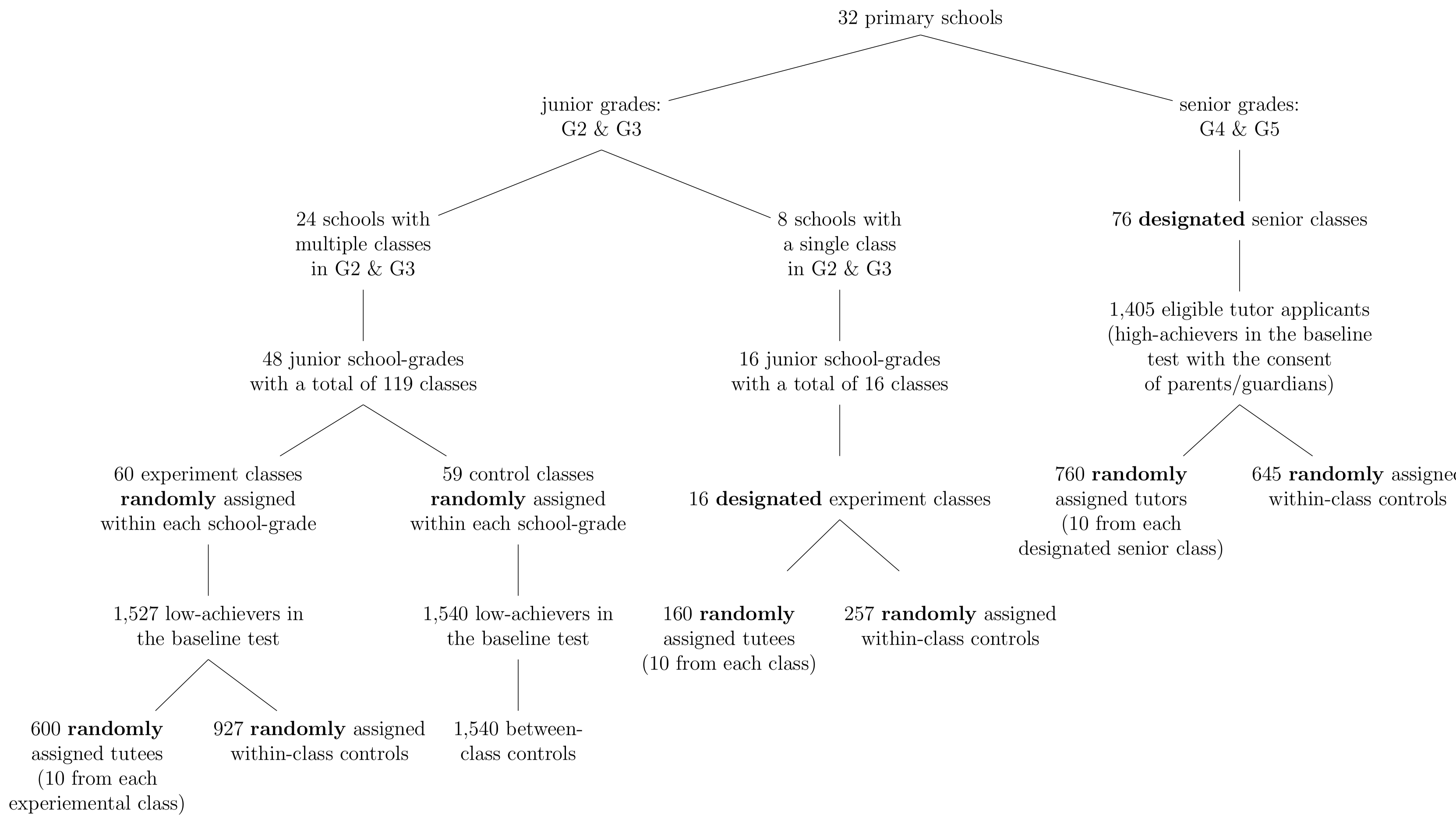

Figure 2: Experiment Design Chart

Table 1: Summary Statistics of Pre-experiment Variables and Sample Balance Checks

Variables	Full sample		Subsample of schools with multiple classes		
	Tutees' mean	Tutees vs. within-class controls diff.	Tutee's mean	Tutees vs. within-class controls diff.	Tutees vs. between-class control diff.
	(1)	(2)	(3)	(4)	(5)
Living w/ one parent only	0.3013	0.0120	0.3150	0.0367	0.0128
	(0.4591)	(0.0215)	(0.4649)	(0.0243)	(0.0223)
Living w/ no parent	0.4368	-0.0232	0.4183	-0.0349	-0.0382
	(0.4963)	(0.0227)	(0.4937)	(0.0255)	(0.0234)
Fraction receiving tutoring	0.3026	-0.0035	0.2800	-0.0052	-0.0369*
at home	(0.4597)	(0.0190)	(0.4494)	(0.0222)	(0.0214)
Average weekly home-tutoring	233.42	-9.97	258.23	-14.36	23.29
time in minutes (excluding 0)	(151.38)	(13.86)	(151.57)	(17.32)	(15.07)
Average weekly home-tutoring	70.64	-3.76	72.30	-4.35	-1.56
time in minutes (including 0)	(135.75)	(6.48)	(140.96)	(7.81)	(6.71)
Baseline math score	-0.7008	-0.0259	-0.6604	-0.0250	-0.0583
	(0.9459)	(0.0366)	(0.9584)	(0.0416)	(0.0377)
Baseline reading score	-0.6910	-0.0120	-0.7152	-0.0375	-0.0536
	(0.9459)	(0.0373)	(0.9527)	(0.0416)	(0.0391)
Number of classes	76	76	60	60	119
Number of students	760	1944	600	1527	2140

Notes:
[1] The full sample consists of all low-achieving grades 2 and 3 students who scored below the class medians in the baseline test in all the 76 experimental classes. Among these students, 10 from each class were randomly assigned as tutees and the remaining were assigned as within-class controls.
[2] The subsample consists of all low-achieving grades 2 and 3 students from a total of 119 classes in 24 primary schools with multiple classes in each grade, from which 60 classes were randomly selected as experimental classes. Among these students, 10 from each class were randomly assigned as tutees, the remaining unselected students from the same experimental classes were assigned as within-class controls, whereas all low-achieving students in the unselected control classes in the same school-grades were assigned as between-class controls.
[3] For each pre-experiment variable denoted by the row heading, columns 1 and 3 report the tutees' mean, columns 2 and 4 the tutees vs. within-class controls difference in means after adjusting for class fixed effects, and column 5 reports the tutee vs. between-class control difference in means after adjusting for school-grade fixed effects. Reported in parentheses are standard deviations for means (columns 1 and 3) and standard errors for differences (columns 2, 4, and 5).
* $p<0.1$

Table 2: Sample Selection Analysis

	Full sample, within-class controls			Subsample, within-class controls			Subsample, btw-class controls		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fraction of tutees with endline scores		0.925			0.935			0.935	
Tutee	-0.0239**	-0.0245**	-0.0199	-0.0157	-0.0167	-0.0193	-0.00760	-0.00884	-0.0114
	(0.0114)	(0.0114)	(0.0152)	(0.0127)	(0.0127)	(0.0167)	(0.0116)	(0.0116)	(0.0152)
Tutee*Living w/ no parent			-0.0105			0.00613			0.00597
			(0.0229)			(0.0255)			(0.0231)
Female		-0.0088	-0.0086		-0.0113	-0.0114		-0.0079	-0.0080
		(0.0114)	(0.0114)		(0.0127)	(0.0128)		(0.0107)	(0.0107)
Living w/ one parent only		-0.0094	-0.0095		-0.0102	-0.0102		-0.0114	-0.0114
		(0.0150)	(0.0150)		(0.0165)	(0.0165)		(0.0138)	(0.0138)
Living w/ no parent		-0.0236*	-0.0196		-0.0224	-0.0247		-0.0291**	-0.0308**
		(0.0142)	(0.0166)		(0.0157)	(0.0184)		(0.0132)	(0.0147)
Class fixed effects	Y	Y	Y	Y	Y	Y			
School-grade fixed effects							Y	Y	Y
Number of classes	76	76	76	60	60	60	119	119	119
Number of observations	1944	1944	1944	1527	1527	1527	2140	2140	2140

Notes: The dependent variable is a dummy indicator for whether a student had taken the endline test. Each column corresponds to a separate regression. Columns 1-3 use tutees and their within-class controls in all the 76 experimental classes; columns 4-6 use tutees and their within-class controls from a subset of 60 experimental classes randomly selected from school-grades with multiple classes; and column 7-9 use tutees from the 60 randomly selected experimental classes and their between-class controls from the unselected classes in the same school-grades. Columns 1-6 control for class fixed effects and columns 7-9 control for school-grade fixed effects. Robust standard errors clustered at the class level are reported in parentheses.

** $p<0.05$, * $p<0.1$

Table 3: Average Treatment Effects on Math and Reading Scores

	Full sample			Subsample, within-class controls			Subsample, btw-class controls		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Math scores									
Tutee	0.149***	0.134***	0.136***	0.102**	0.0904**	0.0939**	0.0867	0.110*	0.110*
	(0.0408)	(0.0395)	(0.0395)	(0.0442)	(0.0434)	(0.0434)	(0.0565)	(0.0569)	(0.0568)
Baseline scores		-0.457***	-0.457***		-0.475***	-0.474***		-0.438***	-0.438***
		(0.0432)	(0.0440)		(0.0518)	(0.0530)		(0.0393)	(0.0396)
Baseline individual controls			Y			Y			Y
Class fixed effects	Y	Y	Y	Y	Y	Y			
School-grade fixed effects							Y	Y	Y
Number of classes	76	76	76	60	60	60	119	119	119
Number of observations	1815	1815	1815	1428	1428	1428	2008	2008	2008
Panel B. Reading scores									
Tutee	0.0148	0.00666	0.0169	0.0143	-0.00311	0.0154	0.0556	0.0316	0.0416
	(0.0422)	(0.0399)	(0.0398)	(0.0498)	(0.0478)	(0.0473)	(0.0628)	(0.0597)	(0.0596)
Baseline scores		-0.315***	-0.343***		-0.333***	-0.363***		-0.324***	-0.350***
		(0.0433)	(0.0448)		(0.0528)	(0.0549)		(0.0321)	(0.0324)
Baseline individual controls			Y			Y			Y
Class fixed effects	Y	Y	Y	Y	Y	Y			
School-grade fixed effects							Y	Y	Y
Number of classes	76	76	76	60	60	60	119	119	119
Number of observations	1815	1815	1815	1428	1428	1428	2008	2008	2008

Notes: The dependent variable is the change in test scores (math for Panel A and reading for Panel B) between the baseline and endline tests. Each column in each panel corresponds to a separate regression. Columns 1-3 use tutees and their within-class controls in all the 76 experimental classes; columns 4-6 use tutees and their within-class controls from a subset of 60 experimental classes randomly selected from school-grades with multiple classes; and column 7-9 use tutees from the 60 randomly selected experimental classes and their between-class controls from the unselected classes in the same school-grades. Columns 1-6 control for class fixed effects and columns 7-9 control for school-grade fixed effects. Columns 3, 6, and 9 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.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Differential Treatment Effects by Parental Absence Status

	Full sample, within-class controls				Subsample, within-class controls		Subsample, btw-class controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline scores	-0.458***	-0.457***	-0.457***	-0.457***	-0.475***	-0.474***	-0.438***	-0.439***
	(0.0434)	(0.0439)	(0.0431)	(0.0439)	(0.0518)	(0.0529)	(0.0393)	(0.0396)
Tutee*Living w/ both parents	0.0908	0.107						
	(0.0682)	(0.0864)						
Tutee*Living w/ one parent only	0.0762	0.0835						
	(0.0517)	(0.0560)						
Tutee*Living w/ no parent	0.203***	0.190***	0.203***	0.190***	0.149***	0.122**	0.134*	0.126*
	(0.0460)	(0.0487)	(0.0461)	(0.0487)	(0.0518)	(0.0528)	(0.0676)	(0.0732)
Tutee*Living with one or both parents			0.0829*	0.0943*	0.0505	0.0735	0.0930	0.0981
			(0.0466)	(0.0530)	(0.0523)	(0.0606)	(0.0641)	(0.0671)
Baseline individual controls		Y		Y		Y		Y
Class fixed effects	Y	Y	Y	Y	Y	Y		
School-grade fixed effects							Y	Y
Number of classes	76	76	76	76	60	60	119	119
Number of observations	1815	1815	1815	1815	1428	1428	2008	2008

Notes: The dependent variable is the change in math scores between the baseline and endline tests. Each column corresponds to a separate regression. Columns 1-4 use tutees and their within-class controls in all the 76 experimental classes; columns 5-6 use tutees and their within-class controls from a subset of 60 experimental classes randomly selected from school-grades with multiple classes; and column 7-8 use tutees from the 60 randomly selected experimental classes and their between-class controls from the unselected classes in the same school-grades. Columns 1-6 control for class fixed effects and columns 7-8 control for school-grade fixed effects. The even columns 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.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Robust Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Baseline scores	-0.448***	-0.450***	-0.457***	-0.457***	-0.449***	-0.450***
	(0.0497)	(0.0496)	(0.0440)	(0.0439)	(0.0496)	(0.0495)
Tutee	0.122**		0.123**		0.110*	
	(0.0483)		(0.0485)		(0.0587)	
Tutee*baseline scores	-0.0207	-0.0171			-0.0200	-0.0166
	(0.0469)	(0.0464)			(0.0466)	(0.0461)
Tutee*female			0.0343	0.0285	0.0327	0.0274
			(0.0749)	(0.0753)	(0.0742)	(0.0746)
Tutee*Living with		0.177***		0.178***		0.167**
no parent		(0.0556)		(0.0577)		(0.0668)
Tutee*Living with		0.0841		0.0843		0.0749
one or both parents		(0.0595)		(0.0588)		(0.0668)
Baseline individual controls	Y	Y	Y	Y	Y	Y
Class fixed effects	Y	Y	Y	Y	N	N
Number of classes	76	76	76	76	76	76
Number of observations	1815	1815	1815	1815	1815	1815

Notes: The dependent variable is the change in math scores between the baseline and endline tests. The sample consists of tutees and their within-class controls in all the 76 experimental classes. Each column corresponds to a separate regression. All regressions control for a female dummy, a dummy indicator for living with only one parent in the baseline, a dummy indicator for living with no parent in the baseline, and class fixed effects. Robust standard errors clustered at the class level are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Bounds on the Average and Differential Treamtne Effects

	Average Treatment Effects			Differential Treatment Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Lower-bound estimations						
Tutee	0.110***	0.101**	0.104***			
	(0.0399)	(0.0387)	(0.0387)			
Tutee*Living w/ no parent				0.178***	0.170***	0.150***
				(0.0518)	(0.0465)	(0.0486)
Tutee*Living w/ one or both parents				0.0611	0.0512	0.0684
				(0.0491)	(0.0451)	(0.0521)
Baseline scores		-0.444***	-0.443***		-0.444***	-0.443***
		(0.0435)	(0.0442)		(0.0434)	(0.0442)
Baseline individual controls			Y			Y
Class fixed effects	Y	Y	Y	Y	Y	Y
Number of classes	76	76	76	76	76	76
Number of observations	1785	1785	1785	1785	1785	1785
Panel B. Upper-bound estimations						
Tutee	0.191***	0.167***	0.169***			
	(0.0429)	(0.0410)	(0.0409)			
Tutee*both parents absent				0.263***	0.239***	0.236***
				(0.0520)	(0.0472)	(0.0513)
Tutee*at least one parent present				0.139**	0.114**	0.118**
				(0.0536)	(0.0483)	(0.0537)
Baseline scores		-0.432***	-0.431***		-0.432***	-0.431***
		(0.0438)	(0.0447)		(0.0437)	(0.0446)
Baseline individual controls			Y			Y
Class fixed effects	Y	Y	Y	Y	Y	Y
Number of classes	76	76	76	76	76	76
Number of observations	1785	1785	1785	1785	1785	1785

Notes: The dependent variable is the change in math scores between the baseline and endline tests. In Panel A, the sample consists of all tutees from the 76 experiment classes and a trimmed subset of their within-class controls excluding the bottom-ranked student in test score changes of the 30 randomly selected classes (one from each class). In Panel B, the sample consists of all tutees from the 76 experiment classes and a trimmed subset of their within-class controls excluding the top-ranked student in test score changes of the 30 randomly selected classes (one from each class). Each column in each panel corresponds to a separate regression. All regressions control for class fixed effects. The regressions in columns 3 and 5 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.
*** $p<0.01$, ** $p<0.05$

Table 7: Treatment Effects on Tutors

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Tutor	0.0169	0.0175	0.0291	0.00334	0.00214	0.0326
	(0.0312)	(0.0289)	(0.0515)	(0.0284)	(0.0257)	(0.0515)
Baseline scores		-0.468***	-0.463***		-0.421***	-0.381***
		(0.0880)	(0.100)		(0.0333)	(0.0460)
Female		-0.109***	-0.103*		0.0770**	0.0558
		(0.0396)	(0.0545)		(0.0309)	(0.0458)
Tutor*baseline scores			-0.00977			-0.0741
			(0.0530)			(0.0474)
Tutor*female			-0.0106			0.0408
			(0.0612)			(0.0536)
Class fixed effects	Y	Y	Y	Y	Y	Y
Number of classes	76	76	76	76	76	76
Number of observations	1372	1372	1372	1372	1372	1372

Notes: The dependent variable is the change in test scores between the baseline and endline tests. The sample consists of all eligible tutor applicants from 76 senior classes. Each column corresponds to a separate regression. All regressions control for class fixed effects. Robust standard errors clustered at the class level are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Self-reported Home Tutoring in the Baseline and Parental Absence Status

	Dummy indicator for home tutoring		Home tutoring time (mins/wk)	
	(1)	(2)	(3)	(4)
(i) Living w/ both parents	0.357***	-	92.2***	-
	(0.019)		(6.4)	
(ii) Living w/ one parent only	0.356***	-	89.6***	-
	(0.018)		(5.9)	
(iii) Living w/ one or both parents	-	0.356***	-	90.8***
		(0.013)		(4.3)
(iv) Living w/ no parent	0.271***	0.271***	61.1***	61.1***
	(0.015)	(0.015)	(4.9)	(4.9)
Difference by the absence of both parents (i.e., (iii) - (iv))	-	0.085***	-	29.7***
		(0.020)		(6.8)
Class fixed effects	Y	Y	Y	Y
Number of classes	76	76	76	76
Number of students	1815	1815	1815	1815

Notes: The dependent variable is a dummy indicator for having reported any tutoring at home in the baseline survey for columns 1-2 and the reported total home tutoring time (in minutes) in the week prior to the baseline survey. The sample consists of tutees and their within-class controls in all the 76 experimental classes. Each column corresponds to a separate regression. All regressions control for class fixed effects. Robust standard errors are reported in parentheses.
*** p<0.01

Table 9: Changes in Self-reported Home-tutoring Inputs

	Tutees	Controls	Combined sample	
			Tutees	Controls
	(1)	(2)	(3)	(4)
Panel A. Change in the dummy indicator for self-reported home tutoring				
(A1) Living w/ one or both parents	-0.164***	-0.170***	-0.153***	-0.168***
	(0.024)	(0.195)	(0.024)	(0.019)
(A2) Living w/ no parent	-0.033	-0.144***	-0.060**	-0.139***
	(0.028)	(0.022)	(0.028)	(0.021)
Difference by the absence of both parents	-0.132***	-0.025	-0.092**	-0.029
(i.e., (A1) - (A2))	(0.039)	(0.030)	(0.037)	(0.029)
Class fixed effects	Y	Y	Y	
Number of classes	76	76	76	
Number of students	703	1112	1815	
Panel B. Change in self-reported home-tutoring time (mins/wk)				
(B1) Living w/ one or both parents	-30.3***	-40.4***	-30.6***	-38.8***
	(8.1)	(7.3)	(8.6)	(7.0)
(B2) Living w/ no parent	-2.5	-33.6***	-10.50	-30.7***
	(9.6)	(8.2)	(10.1)	(7.8)
Difference by the absence of both parents	-27.8**	-6.8	-20.0	-8.00
(i.e., (B1) - (B2))	(13.3)	(11.3)	(13.5)	(10.6)
Class fixed effects	Y	Y	Y	
Number of classes	76	76	76	
Number of students	703	1112	1815	

Notes: The dependent variable is the change in the dummy indicator for having reported any tutoring at home between the baseline and endline surveys in Panel A, and the change in the reported total home tutoring time (minutes per week) between the baseline and endline surveys. Column 1 uses only the tutees subsample and column 2 uses only the controls subsample from 76 experimental classes. Columns 3 and 4 report the results of a single regression using the combined sample of tutees and controls from 76 experimental classes. The coefficients reported in column 3 and 4 are the coefficients on the interaction term between the baseline parental absence status indicated by the row heading and the treatment status indicated by the column heading. All regressions control for class fixed effects. Robust standard errors are reported in parentheses.
*** $p<0.01$, ** $p<0.05$