Received: 26 August 2019 | Revised: 4 April 2020 | Accepted: 16 June 2020

DOI: 10.1002/app5.310

ORIGINAL ARTICLE

ASIA & THE PACIFIC POLICY STUDIES

Does private tutoring improve student learning in China? Evidence from the China Education Panel Survey

Yuhe Guo | **Qihui Chen** | **Shengying Zhai** | **Chunchen Pei**

College of Economics and Management, China Agricultural University, Beijing, China

Correspondence
Qihui Chen, College of Economics and Management, China Agricultural University, Room 302, Beijing 100083, China.
Email: chen1006@umn.edu

Funding information
Humanities and Social Sciences Fund of Ministry of Education of China, Grant/Award Number: 18YJC790010

Abstract

Based on data from the China Education Panel Survey, which covers 28 counties/districts of China, this study applies a difference-in-differences method (combined with propensity score matching in some analyses) to estimate the impacts of private tutoring on students' learning outcomes. Our analyses yield three important findings. First, subject-specific tutoring has a statistically significant and positive effect on Grade 8 students' scores on Chinese and mathematics tests, although the effects are modest in size. Second, private tutoring improves students' academic performance mainly through enhancing their test-taking skills or deepening their understanding of subject-specific knowledge, rather than improving their general cognitive skills. Finally, the effect of private tutoring is heterogenous across different subsamples: it is larger for female students, low-performing students, and students with better-educated and wealthier parents.

KEYWORDS

academic performance, China, education, private tutoring, secondary school

1 | INTRODUCTION

Due to various deficiencies in the formal education system, such as large class sizes and low public educational spending, private tutoring (sometimes referred to as 'shadow education') has become a significant phenomenon in many East Asian and Southeast Asian economies such as China (Zhang, 2014; Zhang & Bray, 2016), Hong Kong (Bray, 2013; Bray, Zhan, Lykins, Wang, & Kwo, 2014), Japan (Sato, 2012; Stevenson & Baker, 1992), Malaysia (Tan & Gibson, 2017), Singapore (Cheo & Quah, 2005), South Korea (Choi & Cho, 2016; Ha & Park, 2017; Ryu & Kang, 2013), Taiwan (Kuan, 2011) and Vietnam (Dang, 2007; Ha & Harpham, 2005). Yet the role private tutoring plays in the development of a country's education system is being hotly debated. Proponents believe that private tutoring provides an effective supplement to the formal education system for both students and tutors (Bhorkar & Bray, 2018; Kim & Park, 2012). This is because, they argue, private tutoring can enhance student learning while simultaneously providing more earning opportunities for regular teachers and (non-teacher track) college students to work as private tutors. Opponents, however, argue that despite these benefits, private tutoring brings about negative side effects at various levels. At the individual level, private tutoring may be inefficient if it crowds out students' attention and energy devoted to learning in regular classes (Cheo & Quah, 2005). At the societal level, the high costs associated with private tutoring may enlarge existing socioeconomic inequality by creating unequal access to educational resources among families with different socio-economic backgrounds (Bray & Kwok, 2003; Damayanthi, 2018).

China provides an interesting case to study. The past decade has witnessed the mushrooming of tutoring institutions in China. In 2016, there were more than 380,000 tutoring institutions nationwide, employing some 8.5 million tutors (Chinese Society of Education, 2016). Tutoring was delivered in a variety of forms, ranging from one-to-one tutoring, to small group meetings, to large-scale lectures, with instructional materials tailored to fit clients' learning needs (Zhang, 2014; Zhang & Bray, 2016). However, observers in China have expressed concerns over this rapidly emerging market, especially on the backwash it has on regular schooling.

A significant proportion (15.9%) of private tutors in China are regular school teachers (Chinese Society of Education, 2016).[1] Providing private tutoring may distract these teachers from their regular teaching duties. Reportedly, some teachers reserve their energy for private lessons, knowing that they will still receive their standard salaries as long as their school-based work is not seriously problematic. Some were even found to have deliberately withheld content in regular classes as a way to 'force' their students to enrol in their private tutoring classes (Zhang, 2014; Zhang & Bray, 2016). Moreover, students may pay more attention in private tutoring sessions (whether or not delivered by their regular school teachers), because they often cover materials that are ahead of the school schedule and/or more exam-oriented, which renders teaching in regular classes more challenging (Liu, 2018). Finally, tutoring–school partnerships in admissions, in which formal schools enrol high-performing students on tests organised by tutoring institutions (the contents of which are usually beyond standard curricula), may distort the official admission procedures and undermine the government's efforts in promoting balanced development of education (Zhang & Bray, 2017). These

[1]Other types of private tutors in China include professional tutors working in tutoring institutions (69.7%), and those in other types of training institutions (9.7%) and college students (5%; Chinese Society of Education, 2016).

problems have led the Chinese Government to issue a series of policies to regulate the private tutoring market.[2]

Yet despite the series of strict regulations recently imposed on its private tutoring market, more than 100 million primary and secondary school students still sought out private tutoring services in 2015 (Chinese Society of Education, 2016). The proportion of students participating in private tutoring (including academic and non-academic tutoring) among primary, lower- and upper-secondary school students reached 47.7%, 47.3% and 51.9%, respectively, in 2017 (Wang, 2017). Each year, a considerable sum is spent on private tutoring by Chinese families—in 2017 alone, an average household spent 5,616 yuan (approximately US$832), amounting to 21.6% of its total disposable income, on private tutoring (National Bureau of Statistics of China, 2018). Time investment is also substantial. It has been estimated that primary students spent 6 hours per week and secondary students 15 hours per week on private tutoring, during ordinary term-time and vacation periods (Huang & Wei, 2018).

Yet, are these investments worthwhile? At a minimum, participation in private tutoring is expected to enhance student learning. Unfortunately, despite its high prevalence, few studies have examined the impact of private tutoring on student learning in China. Two studies, conducted by Zhang (2013) and Zhang and Liu (2016), found that overall, private tutoring does not have significantly positive effects on high school students' scores on the National College Entrance Examination (although it has some significant effects for certain subgroups). However, since both studies were conducted in a single Chinese city (Jinan of Shandong province), their findings may not be generalisable to other settings in China. Studies done in other developing countries also provide little conclusive evidence to help predict possible effects of private tutoring in China. While some found strong positive impacts (Buchmann, 2002; Damayanthi, 2018; Dang, 2007; Tansel & Bircan, 2006), others found only moderate or even no impact (Ha & Harpham, 2005; Loyalka & Zakharov, 2016; Suryadarma, Suryahadi, Sumarto, & Rogers, 2006). Clearly, accurately assessing the effectiveness of private tutoring in China calls for more empirical studies, especially ones with more representative data.

In answering this call, this study employs a dataset recently collected through a large-scale survey that covers 28 counties/districts in China, namely the China Education Panel Survey (CEPS), to examine how private tutoring affects student learning. Exploiting the panel structure of the CEPS data, our difference-in-differences (DID) estimation (combined with propensity score matching in some analyses) yields three important findings. First, subject-specific tutoring has a statistically significant and positive effect on secondary-school students' performance on standardised Chinese and mathematics tests, although these effects are modest in size. Second, and perhaps more importantly, private tutoring in China seems to be test-oriented, rather than skill-oriented, in nature. It improves students' academic performance mainly through enhancing their test-taking skills or providing them with more subject-specific knowledge

[2]In 2010, the State Council of China issued the 'National Plan for Medium- and Long-Term Education Reform and Development (2010–2020)', highlighting the heavy burden private tutoring imposes on Chinese students. Strict regulations were then issued on the employees, minimum qualifications required, and the instruction content of private tutoring institutions at the basic-education level (Grades 1–9). In 2015, the Ministry of Education issued a policy prohibiting in-service school teachers from organising or providing private tutoring. In February 2018, the Ministry of Education, Ministry of Civil Affairs, Ministry of Human Resources and Social Security, and the State Administration for Industry and Commerce jointly issued the 'Notice on Reducing the Overburden of Primary and Secondary School Students and Implementing Special Actions for Private Tutoring Institutions' and implemented several measures to regulate the private tutoring market. Subsequently, all 31 provinces/municipalities/autonomous regions in mainland China released their guidelines in April 2018 (Zhang, 2019).

rather than improving their general cognitive skills. Finally, the effect of private tutoring is heterogenous across subgroups with different characteristics, which has profound implications for educational inequality: it is larger for girls and originally low-performing students, which helps to reduce educational inequality, but also larger for students with better-educated and wealthier parents, which may lead to enlarged inequality.

The rest of the article proceeds as follows. The next section describes the CEPS data and develops an analytical framework underlying our empirical analysis. Section 3 presents and discusses our key findings. The final section concludes and discusses some limitations of our study.

2 | METHOD

2.1 | Data

2.1.1 | Survey

The data analysed in this study were drawn from CEPS, a panel survey designed by the Renmin University of China to be a nationally representative survey of lower-secondary school students (Grades 7–9).[3] In the baseline academic year of 2013–2014, the CEPS adopted a multi-stage Probability-Proportional-to-Size sampling method to select sample students. It first used the average level of education and the proportion of migrants as stratifying variables to select 28 rural counties/urban districts from all counties/districts in China. Next, enrolment size and school type (i.e., public, private and migrant schools) were used as stratifying variables to randomly select four lower-secondary schools from each selected county/district.[4] In each of the selected schools, four classes (two from Grade 7 and two from Grade 9) were then randomly chosen; all students enrolled in these classes participated in the survey.[5] A total of 19,487 students in 438 selected classes were interviewed. Four rounds of follow-up surveys were conducted from 2014 to 2018, one in each academic year, but currently only two rounds of data (2013–2014 and 2014–2015) are publicly available. Our analysis is therefore based on these two available rounds of data. In particular, we restrict our attention to (the 10,279) students who were in Grade 7 in the baseline survey (9,449 or 91.93% of whom were followed in Round 2), because the majority of the baseline 9th graders had graduated by the time of Round 2 and cannot be followed. (For reasons that will become clear below, our empirical analyses are further restricted to the 5,733 baseline 7th graders who were interviewed in the first semester starting in September.)

During each survey round, separate questionnaires were administered to the sampled students, their parents, homeroom teachers, teachers of main subjects (e.g., Chinese and mathematics), and school principals to collect relevant information on these students' educational development. The information collected includes students' personal characteristics (e.g., gender,

[3]More details on the CEPS can be found on its official website: https://ceps.ruc.edu.cn/.

[4]In China, ‘migrant schools’ refer to those (usually privately funded) schools that are constructed to temporarily accommodate children of rural migrants working in the cities (Chen & Feng, 2013). While migrant schools adopt the same curriculum as regular schools to fulfil their role as education facilities, according to regulations in the majority of localities, migrant students without legal permanent residential permits must return to their hometowns to pursue high school education (Wang, Luo, Zhang, & Rozelle, 2017).

[5]If there are no more than two classes in a targeted grade in a given school, all students in that grade were interviewed.

age, academic performance and private tutoring participation status), family characteristics (e.g., parental education, family wealth, household configuration, and household members' employment status), subject teachers' qualifications (e.g., educational background and teaching experience), as well as class and school characteristics (e.g., class size and condition of school facilities).

The most important information provided by the schools, which enables us to examine the private tutoring–student learning relationship, is their students' academic performance, measured by scores on Chinese and mathematics mid-term tests taken during a given semester. According to CEPS documentation, these tests were designed by the sampled schools based on their subject-specific syllabi to capture students' knowledge of subject-specific content.[6] For ease of interpretation and comparison, we standardised the original test scores to have zero mean and unit standard deviation (SD) within schools.

Note that the timing of the survey adopted by the CEPS adds a layer of complexity for examining the private tutoring–student learning relationship. Each school involved in the CEPS was visited either in the first (September) or second (March) semester of an academic year, and for a given school, depending on the timing of the first visit, the survey semester in different rounds remained the same (in other words, if a school was visited in the first semester, then it would be visited only in the first semester in subsequent rounds). A problem with such a survey design is that while the CEPS collected information on students' private-tutoring participation *during* the survey semester (either first or second), it collected mid-term exam scores uniformly for only the *first* semester (of the academic year of survey), regardless of the actual survey semester (first or second). To ensure that information on students' private tutoring participation status matches that on their test scores recorded in the *same* semester, we restrict our main analyses to the sample of 5,733 then 7th graders who were interviewed in the first semester (but not those interviewed in the second semester) during the baseline survey.[7]

2.1.2 | Sample characteristics

In this restricted sample, slightly more than half (52%) are boys. The average student was 13 years old in 2014, and came from a three- or four-membered nuclear family, whose father and mother had completed 10.5 and 9.9 years of education, respectively. This student was enrolled in a class size of 51, was being taught by a Chinese teacher with 15.5 years of education and 15.7 years of teaching experience, and a maths teacher with 15.4 years of education and 16.2 years of experience.

Among all 5,733 students in the analytical sample, 1,776 (31%) participated in maths tutoring and 1,007 (18%) participated in Chinese tutoring sometime between the baseline and the Round 2 surveys. As suggested in panel A of Table 1, which presents descriptive statistics for these 5,733 students (pooling data from both rounds), (subject-specific) private tutoring participants outperformed non-participants on both Chinese and mathematics tests (although

[6] China Education Panel Survey, https://ceps.ruc.edu.cn/index.php?r=index/index&hl=en (accessed on May 20, 2019).

[7] Note that this restriction is unlikely to introduce sample-selection problems because the decision to visit a school in the first (September) or the second (March) semester is made by the CEPS project team, mostly based on logistics considerations, which are exogenous to students' and schools' behaviour.

TABLE 1 Summary statistics, by private tutoring participation status, all 7th graders interviewed in the first semester at the baseline

	Chinese tutoring					Maths tutoring				
	Non-participants		Participants		Difference	Non-participants		Participants		Difference
Variables	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Panel A: Outcome variables										
Chinese test scores (standardised)	−0.08	[0.96]	−0.01	[0.88]	0.07**	−0.09	[0.97]	−0.02	[0.87]	0.07***
Maths test scores (standardised)	−0.02	[0.95]	0.10	[0.89]	0.11***	−0.01	[0.96]	0.02	[0.89]	0.03
Panel B: Explanatory/control variables										
Age (months)	158.76	[9.90]	157.64	[8.69]	−1.12***	158.67	[10.09]	158.53	[8.48]	−0.14
Male (=1 if yes)	0.52	[0.50]	0.53	[0.50]	0.01	0.53	[0.5]	0.46	[0.50]	−0.07***
Cognitive ability test score (standardised)	0.12	[0.81]	0.30	[0.84]	0.18***	0.10	[0.83]	0.29	[0.74]	0.19***
Birth order	1.35	[0.65]	1.21	[0.52]	−0.14***	1.37	[0.67]	1.20	[0.50]	−0.17***
Number of siblings	0.72	[0.83]	0.52	[0.74]	−0.20***	0.75	[0.83]	0.49	[0.75]	−0.26***
Father's education (years)	10.32	[3.0]	11.80	[3.26]	1.48***	10.15	[2.99]	11.73	[3.22]	1.58***
Mother's education (years)	9.72	[3.32]	11.28	[3.53]	1.56***	9.53	[3.27]	11.31	[3.41]	1.78***
Household registration (=1 if rural)	0.57	[0.50]	0.39	[0.49]	−0.18***	0.60	[0.49]	0.36	[0.48]	−0.24***
Income-level dummies:										
Very poor	0.04	[0.21]	0.02	[0.13]	−0.02***	0.05	[0.21]	0.01	[0.11]	−0.04***
Poor	0.20	[0.40]	0.13	[0.34]	−0.07***	0.21	[0.41]	0.12	[0.32]	−0.09***
On average	0.71	[0.45]	0.76	[0.43]	0.05***	0.70	[0.46]	0.79	[0.41]	0.09***
Rich	0.04	[0.21]	0.09	[0.28]	0.05***	0.04	[0.20]	0.08	[0.26]	0.04***
Very rich	0.00	[0.05]	0.01	[0.09]	0.01***	0.00	[0.05]	0.00	[0.07]	0.00
Class size	51.12	[13.60]	49.17	[13.44]	−1.95***	51.09	[13.6]	50.22	[13.57]	−0.87***
Subject teacher's education (years)	15.49	[0.80]	15.58	[0.85]	0.09***	15.38	[1.09]	15.64	[0.77]	0.26***
Subject teacher's teaching experience (years)	15.62	[8.81]	15.89	[9.40]	0.27	16.21	[8.36]	16.14	[8.44]	−0.07
Number of students-by-round observations	10,254		1,212			9,143		2,323		

Note: The sample involves only the 5,733 students who were 7th graders and were interviewed in the first (September) semester at the baseline in 2014. Information collected in both rounds of survey (baseline and Round 2 survey) is used to compute the descriptive statistics. Standard deviations (SD) in brackets.
$^{**}p < 0.01$, $^{***}p < 0.001$.

the differences are not always statistically significant). However, participants and non-participants also differ significantly in many of their personal and family characteristics (Table 1, panel B); thus, the private tutoring–student learning associations observed in panel A of Table 1 may simply reflect the influences of these (and other, unobserved) characteristics. To further infer a causal effect of private tutoring on better academic performance, more careful investigation is needed.

2.1.3 | Dynamics of private tutoring participation

A closer investigation into the dynamics of private tutoring in the two rounds of data reveals non-trivial 'take-up' and 'drop-out' behaviours. More specifically, among the 5,733 students in the analytical sample, 395 (7%) did not participate in Chinese tutoring at the baseline but participated in it at the time of the Round 2 survey; in contrast, 407 (7%) baseline participants tutored in Chinese had quit Chinese tutoring by Round 2. Meanwhile, whereas 824 (14%) baseline non-participants of maths tutoring became participants in Round 2, 405 (7%) baseline participants dropped out of maths tutoring in Round 2. The remaining students did not change their participation status between the two rounds of surveys: 4,726 (82%) never participated in Chinese tutoring and 3,957 (69%) never participated in maths tutoring during either round; 205 (4%) and 547 (10%) students participated in Chinese and maths tutoring, respectively, in both rounds.

These patterns naturally divide the sample into two groups, the 'take-up' group, which consists of students who did not participate in private tutoring at the baseline (among whom some began to participate in Round 2), and the 'drop-out' group, which consists of students who had already participated in private tutoring at the baseline (among whom some dropped out in Round 2). Such a group division allows us to examine the private tutoring–learning relationship more deeply.

Note that the dynamics of private tutoring participation allows us to perform some preliminary DID (difference-in-differences) analyses for both the 'take-up' and 'drop-out' groups. Simple DID calculations presented in Table 2 show strong positive correlations between private tutoring participation and academic performance. Consider Chinese tutoring in the 'take-up' group (Table 2, panel A, columns 1–3), for example. In the 'take-up' group, no one participated in private tutoring at the baseline. Yet in Round 2, while the Chinese test scores of those who remained as non-participants increased by 0.167 SDs between the two rounds of survey, the scores of those participating in Round 2 increased by as much as 0.372 SDs, suggesting a sizeable positive impact (0.205 SDs = 0.372–0.167 SDs) of Chinese tutoring. A similar pattern is revealed in the 'drop-out' group: students in this group all participated in private tutoring at the baseline; but whereas the Chinese scores of those who remained as participants increased by 0.321 SDs over time, the scores of those who had quit Chinese tutoring by Round 2 increased by only 0.114 SDs. Similar patterns are also seen in maths tutoring (Table 2, panel B), although the contrasts are less impressive.

It should be kept in mind that these simple DID estimates are only suggestive of the positive impacts of private tutoring, in that these estimates were obtained without controlling for potential confounding factors. Also, as will be shown in the next section, only the DID estimates defined over the 'take-up' group (but not those defined over the 'drop-out' group) provide sensible estimates of meaningful underlying population parameters.

TABLE 2 Dynamics of private tutoring participation and academic performance

Groups	'Take-up' group			'Drop-out' group		
	'Non-participants' at the baseline			'Participants' at the baseline		
	(1) 'Non-participants' in Round 2	(2) 'Participants' in Round 2	(3) Difference = (2)–(1)	(4) 'Non-participants' in Round 2	(5) 'Participants' in Round 2	(6) Difference = (5)–(4)
Panel A. Standardised Chinese test scores						
(a) Baseline	−0.162 [0.931]	−0.184 [0.789]	−0.022 (0.050)	−0.200 [0.967]	−0.176 [0.792]	0.024 (0.079)
(b) Round 2	0.005 [0.996]	0.188 [0.793]	0.183^{***} (0.051)	−0.086 [0.983]	0.145 [0.85]	0.231^{***} (0.081)
(c) Difference = (b) – (a)	0.167^{***} (0.020)	0.372^{***} (0.056)		0.114^{*} (0.065)	0.321^{***} (0.081)	
(d) Difference in differences			0.205^{***} (0.070)			0.207^{*} (0.112)
Number of observations	4,726	395		407	205	
Panel B. Standardised maths test scores						
(e) Baseline	−0.015 [0.939]	0.006 [0.785]	0.021 (0.036)	−0.106 [0.916]	−0.079 [0.807]	0.027 (0.060)
(f) Round 2	0.003 [0.999]	0.137 [0.88]	0.134^{***} (0.038)	−0.092 [0.998]	0.029 [0.949]	0.121^{*} (0.064)
(g) Difference = (f) – (e)	0.018 (0.021)	0.131^{***} (0.041)		0.014 (0.064)	0.108^{*} (0.053)	
(h) Difference in differences			0.113^{**} (0.051)			0.094 (0.085)
Number of observations	3,957	824		405	547	

Note: The analytical sample includes only students who were 7th graders and were interviewed in the first (September) semester at the baseline in 2014. Two rounds (baseline and Round 2) of data are used. The 'take-up' group consists of students who did not participate in private tutoring at the baseline (some of whom participated in Round 2); the 'drop-out' group consists of students who participated at the baseline (some of whom dropped out in Round 2). Standard deviations in brackets; standard errors in parentheses.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

2.2 | Difference-in-differences estimation

2.2.1 | Identifiable parameters

With variations in students' private tutoring participation status over time, a natural estimator of the impacts of private tutoring is the DID estimator. However, with the concurrent occurrence of 'take-up' and 'drop-out' behaviours, the 'treatment group' required in a standard impact evaluation study is not straightforward to define. It is therefore helpful to formally derive what can really be identified with available data.

Let a binary indictor $P = P(t)$ denote a student's private tutoring participation status ($P = 1$ for participants and $P = 0$ for non-participants at time t), where t represents the time period of observation ($t = t_0$ for the baseline and $t = t_1$ for the follow-up, Round 2 period). The sample of students can then be subdivided into four subgroups according to their participation status in the two periods: $[P(t_0), P(t_1)] = [(0, 0), (0, 1); (1,1), (1,0)]$. The first two subgroups (i.e., $[P(t_0), P(t_1)] = (0, 0)$ or $(0, 1)$) combine to form the 'take-up' group (in which no one participated in private tutoring at the baseline, but some participated in Round 2); the latter two subgroups (i.e., $[P(t_0), P(t_1)] = (1, 1)$ or $(1, 0)]$ combine to form the 'drop-out' group (in which all students participated in private tutoring at the baseline but some had quit by Round 2).

Adopting the potential outcome framework for causal inference (Rubin, 1974), let Y^1 and Y^0 denote a student's treated and untreated outcomes, respectively. For reasons that will soon be clear, our focus is first on the 'take-up' group. A simple cross-section (CS) comparison of an observed outcome Y (say, Chinese test scores) between participants ($P = 1$) and non-participants ($P = 0$) in the follow-up period (t_1) is

$$
\begin{aligned}
\beta^{CS} &= E(Y|P=1,\ t_1) - E(Y|P=0,\ t_1) \\
&= E(Y^1|P=1,\ t_1) - E(Y^0|P=0,\ t_1) \\
&= [E(Y^1|P=1,\ t_1) - E(Y^0|P=1,\ t_1)] \quad + [E(Y^0|P=1,\ t_1) - E(Y^0|P=0,\ t_1)] \\
&= ATT^{take-up} \quad + [E(Y^0|P=1,\ t_1) - E(Y^0|P=0,\ t_1)],
\end{aligned}
\tag{1}
$$

where $ATT^{take-up} = E(Y^1|P = 1,\ t_1) - E(Y^0|P = 1,\ t_1)$ is the *average treatment effect on the treated* (ATT) *defined over the 'take-up' group*; the term $[E(Y^0|P = 1,\ t_1) - E(Y^0|P = 0,\ t_1)]$ is the so-called 'selection bias', more specifically, the participant–non-participant difference in the *untreated* outcome (Y^0) at t_1.

Under the (identifying) assumption that the participant–non-participant difference in the untreated outcome is fixed over time (the so-called 'parallel trends' assumption),[8] that is $[E(Y^0|P = 1,\ t_1) - E(Y^0|P = 0,\ t_1)] = [E(Y^0|P = 1,\ t_0) - E(Y^0|P = 0,\ t_0)]$, which implies that $[E(Y^0|P = 1,\ t_1) - E(Y^0|P = 0,\ t_1)] = [E(Y|P = 1,\ t_0) - E(Y|P = 0,\ t_0)]$, an unbiased estimate of $ATT^{take-up}$ can be obtained by subtracting the observed participant–non-participant difference at the baseline, $[E(Y|P = 1,\ t_0) - E(Y|P = 0,\ t_0)]$, from β^{CS} defined in Equation (1):

[8]Note that $[E(Y^0|P = 1,\ t_1) - E(Y^0|P = 0,\ t_1)] = [E(Y^0|P = 1,\ t_0) - E(Y^0|P = 0,\ t_0)]$ is equivalent to $[E(Y^0|P = 1,\ t_1) - E(Y^0|P = 1,\ t_0)] = [E(Y^0|P = 0,\ t_1) - E(Y^0|P = 0,\ t_0)]$. In words, *in the absence of treatment* (private tutoring), the changes in Y between t_0 and t_1 are *identical* in the participant ($P = 1$) and non-participant ($P = 0$) groups – in other words, the time trends in untreated outcomes (Y^0) are *parallel* in the two groups.

$$
\begin{aligned}
ATT^{take-up} &= \beta^{CS} - [E(Y1vt_0) - E(Y0vt_0)] \\
&= [E(Y1vt_1) - E(Y0vt_1)] - [E(Y1vt_0) - E(Y0vt_0)] \\
&= [E(Y1vt_1) - E(Y1vt_0)] - [E(Y0vt_1) - E(Y0vt_0)],
\end{aligned} \tag{2}
$$

which is a DID estimator defined over the 'take-up' group (note that since the 'take-up' group consists of all students who did not participate in private tutoring at the baseline, the underlying population of this group is also well-defined).

Note, however, that a similar derivation reveals that ATT defined over the '*drop-out*' group, $ATT^{drop\text{-}out}$, cannot be estimated with available data. This is because the participant–non-participant difference in the observed outcome at the baseline $[E(Y|\, P=1,\ t_0) - E(Y|\, P=0,\ t_0)]$ is the difference in *treated* outcomes (i.e., $E(Y^1|\, P=1,\ t_0) - E(Y^1|\, P=0,\ t_0)$), and it is unlikely that this difference will be able to cancel out the selection-bias term $[E(Y^0|\, P=1,\ t_1) - E(Y^0|\, P=0,\ t_1)]$ embedded in β^{CS} in Equation (1), especially when private tutoring is indeed effective at enhancing learning (so that $Y^1 > Y^0$).

2.2.2 | Estimation strategy

It can be easily shown that the DID estimate of $ATT^{take\,-\,up}$ can be estimated by the following linear regression model applied to the 'take-up' group:

$$Y = \beta_0 + \beta_1 P + \beta_2 T + \beta^{ATT}(P \times T) + \varepsilon, \tag{3}$$

where ε is the disturbance term that satisfies $E[\varepsilon|P,\ T] = 0$. Equivalently, the first-differenced model

$$\Delta Y = \gamma_0 + \beta^{ATT}\Delta P + v \tag{4}$$

can be estimated, where ΔY and ΔP are the changes in the values of Y and P between the baseline and Round 2.

Recall that the identifying assumption (the 'parallel trends' assumption) needed for this DID estimator to yield an unbiased estimate of $ATT^{take\,-\,up}$ is $[E(Y^0|\, P=1,\ t_1) - E(Y^0|\, P=0,\ t_1)] = [E(Y|\, P=1,\ t_0) - E(Y|\, P=0,\ t_0)]$. With data collected from only two time periods, this assumption is essentially *untestable*. Adopting a second-best alternative, we relax this assumption to be $[E(Y^0|\, P=1,\ t_1,\ \mathbf{X}) - E(Y^0|\, P=0,\ t_1,\ \mathbf{X})] = [E(Y|\, P=1,\ t_0,\ \mathbf{X}) - E(Y|\, P=0,\ t_0,\ \mathbf{X})]$, where $\mathbf{X}$ is a set of personal, family and class/school characteristics (Table 1, panel B). In words, the relaxed identifying assumption says that the participant–non-participant difference in untreated outcome is fixed over time, provided that a set of observed (baseline) characteristics $\mathbf{X}$ has been conditioned on. Under this relaxed assumption, the modified DID estimator is:

$$Y = \beta_0 + \beta_1 P + \beta_2 T + \beta^{ATT}(P \times T) + \mathbf{X}\boldsymbol{\beta} + \varepsilon. \tag{5}$$

Another way to condition the DID estimator on $\mathbf{X}$ is to first perform propensity score matching using a set of $\mathbf{X}$ variables observed at the baseline, and then apply DID to the matched sample over the common support C:

$$Y = \beta_0 + \beta_1 P + \beta_2 T + \beta^{ATT}(P \times T) + \varepsilon | Pr(\mathbf{X}) \in \mathrm{C}, \tag{6}$$

or

$$\Delta Y = \gamma_0 + \beta^{ATT} \Delta P + v | Pr(\mathbf{X}) \in \mathrm{C}, \tag{7}$$

where $Pr(\mathbf{X})$ is the estimated propensity scores (i.e., the estimated probabilities of participating in subject-specific private tutoring in Round 2).[9] To assess the robustness of our estimates, we use both conditioning strategies in the analyses reported and discussed in the next section.

3 | RESULTS

3.1 | Effects of private tutoring on academic performance

Turning to the main findings of this article, Table 3 reports the results of estimating the impacts of private tutoring, for standardised Chinese (columns 1–3) and maths (columns 4–6) test scores. For either outcome, the first column uses subject-specific private tutoring as the only explanatory variable in the model, the second column controls for a large set of baseline covariates, and the third reports results of DID estimation combined with propensity score matching (results of the first-stage propensity scores estimation based on a logit model, and the resulting common support, constructed based on the nearest neighbour algorithm, are reported in Appendix A, Table A1 and Figure A1).

The results show that subject-specific private tutoring has a statistically significant and positive effect on students' performance in the corresponding subject. Specifically, other things being equal, subject-specific tutoring raised students' Chinese and maths test scores by 0.10–0.14 SDs and 0.07–0.09 SDs, respectively. The estimated impacts are quite similar across subjects, empirical specifications and estimators. However, these impacts are relatively small compared to previous findings in some other Asian countries, such as India (Banerjee, Cole, Duflo, & Linden, 2007) and Thailand (Thongphat, 2012), where private tutoring is found to raise students' test scores by more than 0.2 SDs.

The estimated coefficients of the control variables are informative and consistent with previous findings. For example, girls outperformed boys in both subjects, which is commonly observed in lower-secondary schools in China in recent years (e.g., Chen, 2017; Chen, Wang, & Zhao, 2019; Lai, 2010). Also intuitively, students' cognitive ability scores (discussed below) and parental education positively predict their achievement test scores, although not always in a statistically significant manner.

3.2 | Potential channels

For both academic and policy purposes, it is helpful to understand the channels through which private tutoring improves students' academic performance. A number of channels might be at

[9] In our propensity score matching analysis, a logit model is used to estimate the propensity scores; estimation results are reported in Appendix A, Table A1. Matching is then based on the nearest neighbour algorithm; the resulting common support region is presented in Figure A1.

TABLE 3 Impacts of subject-specific private tutoring on standardised test scores

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome variable	Chinese test scores (standardised)			Maths test scores standardised)		
Estimator	DID	DID	DID-PSM	DID	DID	DID-PSM
Subject-specific tutoring	0.107*** (0.036)	0.099*** (0.036)	0.135** (0.057)	0.081** (0.034)	0.071** (0.032)	0.087*** (0.033)
Child-level controls						
Age (months)		−0.002 (0.002)			−0.003* (0.002)	
Boy (=1 if yes)		−0.146*** (0.022)			−0.084*** (0.023)	
Cognitive ability score (standardised)		0.070*** (0.021)			0.060** (0.024)	
Birth order		0.008 (0.023)			−0.011 (0.021)	
Sibship size		−0.017 (0.021)			0.006 (0.019)	
Family-level controls						
Father's education (years)		0.001 (0.006)			0.003 (0.005)	
Mother's education (years)		0.003 (0.005)			0.009* (0.004)	
Household registration (=1 if rural)		−0.012 (0.026)			0.028 (0.026)	
Household income (dummies)		Yes			Yes	
Class/school-level controls						
Class size		0.013 (0.008)			−0.007 (0.007)	
Subject teacher's education (years)		−0.016 (0.054)			−0.043*** (0.011)	
Subject teacher's experience (years)		−0.001 (0.004)			−0.003 (0.003)	

(Continues)

TABLE 3 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
School fixed effects (dummies)	Yes	Yes		Yes	Yes	
Constant	−0.079 (0.067)	0.043 (0.974)	0.237*** (0.075)	−0.254*** (0.041)	1.238** (0.513)	0.215*** (0.068)
R^2	0.253	0.267	0.007	0.212	0.226	0.006
Number of (matched) observations	5,121	5,121	790 (treated: 395; untreated: 395)	4,781	4,781	1,648 (treated: 824; untreated: 824)
Number of schools	68	68	62	68	68	65

Note: The analysis is restricted to the 'take-up' group and the then 7th graders interviewed in the first (September) semesters. Income-level dummies include 'very poor' (baseline category) 'poor', 'on average', 'rich' and 'very rich' (Table 1). Results of the first-stage propensity scores estimation, based on a logit model, and the estimated common support, based on the nearest neighbour algorithm, are reported in Appendix A. Robust standard errors in parentheses, adjusted for within-school clustering.
Abbreviations: DID, difference in differences; PSM, propensity score matching.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

work. First, the performance-improving effect of private tutoring detected above (Table 3) may work through developing more advanced cognitive skills, for example rigorous analytical and logical thinking skills in both general and subject-specific terms, in students—a skill-developing effect. Second, private tutoring may develop other types of skills, say, test-taking skills, in students, with or without improving their cognitive skills. The third effect is a subject-specific knowledge-accumulation effect: private tutoring may deepen students' understanding of subject-specific materials that they did not fully understand in regular classes, say, through repeated practice, without actually improving their cognitive skills.

Although the CEPS data do not contain direct information for capturing all these potential channels, we try to infer their existence by a series of (indirect) tests. The first is to see whether private tutoring has any general (as opposed to subject-specific) skill-improving effect. This is done by adding a dummy variable capturing participation in private tutoring *in other subjects* to the DID model (Equation 5). Yet as shown in Table 4 (columns 1 and 3), the impact of 'tutoring in other subjects' is statistically insignificant; meanwhile, the inclusion of it leads to no notable changes in the estimated coefficients of other variables.[10] Also, excluding dummies for subject-specific private tutoring (Table 4, columns 2 and 4) does not notably improve the predictive power of the other-subject-tutoring dummy for test scores, suggesting that the impact of private tutoring in China is mostly subject-specific.

We then test if private tutoring improves students' academic performance through developing more cognitive skills (as opposed to equipping them with more subject-specific knowledge or improving their test-taking skills). To this end, we estimate the impacts of private tutoring on students' scores on a cognitive ability test (Table 1, panel B), which was designed by the CEPS team and administered in all sampled schools. The content of the test is designed to be *not* closely related to the specific subject knowledge taught in regular classes. And to help ensure that students will not be distracted by the context of the test questions when taking the test, the CEPS team designed test questions based on information and scenarios drawn from

TABLE 4 Impacts of subject-specific private tutoring on standardised test scores

	(1)	(2)	(3)	(4)
Outcome variable	Chinese scores (standardised)		Maths scores (standardised)	
Estimator	DID		DID	
Subject-specific tutoring	0.090^{**} (0.038)		0.059^{*} (0.033)	
Tutoring in other subjects	0.022 (0.027)	0.036 (0.026)	0.032 (0.031)	0.048 (0.031)
Baseline controls	Yes	Yes	Yes	Yes
R^2	0.267	0.266	0.226	0.225
Number of observations	5,121	5,121	4,781	4,781
Number of schools	68	68	68	68

Note: The analysis is restricted to the 'take-up' group and students interviewed in the first (September) semesters. Baseline control variables include the full set reported in Table 3. Robust standard errors in parentheses, adjusted for within-school clustering.
Abbreviation: DID, difference in differences.
$^{*}p < 0.05$, $^{**}p < 0.01$.

[10]Detailed regression results not shown but available on request.

everyday life. Put differently, the cognitive ability test mainly measures students' general logical and analytical thinking skills, rather than subject-specific knowledge and test-taking skills. The main results of this investigation, reported in Table 5, show that neither Chinese tutoring (columns 1–3) nor maths tutoring (columns 4–6) leads to significant improvements in students' general cognitive ability. This implies that private tutoring in China is likely to improve students' academic performance either through the (subject-specific) knowledge-accumulation effect or by improving their (subject-specific) test-taking skills—but unfortunately, the data lack information to distinguish these two channels.

3.3 | Heterogenous effects

More insights into why private tutoring improves student learning in China may be obtained by investigating how the effects of private tutoring vary across different subsamples. Table 6 repeats the DID analyses reported in columns 2 and 5 of Table 3, but this time separately by gender groups (panels A–B), baseline academic performance (panels C–D), parental education (panels E–H), and family wealth (panels I–J).

Three informative patterns emerge from the results reported in Table 6. First, there are gender differences in the effects of private tutoring: compared to boys (panel B), girls (panel A) are more likely to benefit from private tutoring. This may be due to gender differences in learning habits, as girls are usually found to be more self-disciplined than boys (Duckworth & Seligman, 2006). Second, private tutoring is more effective for low-performing students (panel C) than for high-performing ones (panel D), which suggests that private tutoring in China serves a remedial function for students who are left behind. Finally, students with relatively more advantageous family backgrounds, for example those with better-educated parents (panels F and H) and from wealthier families (panel J), are more likely to benefit from private tutoring. Intuitively, better family conditions enable these students to seek more and better private

TABLE 5 Impacts of private tutoring on cognitive ability test scores

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome variables	Cognitive ability test scores (standardised)			Cognitive ability test scores (standardised)		
Type of tutoring	Chinese tutoring			Maths tutoring		
Estimator	DID			DID		
Chinese/maths tutoring	0.007 (0.041)	−0.015 (0.042)		0.031 (0.032)	0.015 (0.033)	
Tutoring in other subjects		0.053 (0.032)	0.050 (0.031)		0.041 (0.033)	0.045 (0.032)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.122	0.123	0.123	0.125	0.126	0.126
Number of observations	5,121	5,121	5,121	4,781	4,781	4,781
Number of schools	68	68	68	68	68	68

Note: The analysis is restricted to the 'take-up' group and students interviewed in the first (September) semesters. Robust standard errors in parentheses, adjusted for within-school clustering. Baseline controls include the full set reported in Table 3.

Abbreviation: DID, difference in differences.

TABLE 6 Heterogenous impacts of private tutoring

	(1)	(2)	(3)	(4)
Test scores (standardised)	Chinese	Maths	Chinese	Maths
	A. Girls		B. Boys	
Subject-specific tutoring	0.108** (0.048)	0.065* (0.039)	0.079 (0.058)	0.068 (0.049)
Baseline controls	Yes	Yes	Yes	Yes
N	2,471	2,261	2,650	2,520
R^2	0.362	0.265	0.216	0.218
	C. Subject score (baseline) < Median		D. Subject score (baseline) ≥ Median	
Subject-specific tutoring	0.152** (0.057)	0.095** (0.045)	0.059 (0.040)	0.040 (0.035)
Baseline controls	Yes	Yes	Yes	Yes
	2,566	2,382	2,555	2,399
R^2	0.214	0.211	0.234	0.233
	E. Father's education < Median		F. Father's education ≥ Median	
Subject-specific tutoring	0.048 (0.056)	0.064 (0.047)	0.114*** (0.042)	0.079** (0.038)
Baseline controls	Yes	Yes	Yes	Yes
N	2,667	2,556	2,454	2,225
R^2	0.233	0.190	0.324	0.282
	G. Mother's education < Median		H. Mother's education ≥ Median	
Subject-specific tutoring	0.077 (0.060)	0.032 (0.038)	0.100** (0.040)	0.122** (0.051)
Baseline controls	Yes	Yes	Yes	Yes
N	2,635	2,525	2,486	2,256
R^2	0.230	0.192	0.325	0.283
	I. Family wealth: poor		J. Family wealth: not poor	
Subject-specific Tutoring	0.015 (0.090)	0.042 (0.076)	0.112*** (0.040)	0.080** (0.035)
Baseline controls	Yes	Yes	Yes	Yes
N	1,260	1,227	3,861	3,554
R^2	0.214	0.196	0.296	0.249

Note: The analysis is restricted to the 'take-up' group and students interviewed in the first (September) semesters. Robust standard errors in parentheses, adjusted for within-school clustering. All regressions include school fixed effects and all control variables reported in Table 3.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

tutoring services (Buchmann, Condron, & Roscigno, 2010), thereby enhancing the effectiveness of private tutoring for these students. These findings suggest that the performance-improving effect of private tutoring might be achieved through its interaction with students' background characteristics.

4 | DISCUSSION

Deficiencies in the education system, such as large class size and low public educational spending, are often cited as the main reasons for the rapid expansion of private tutoring services in developing countries. Private tutoring is also becoming increasingly popular among students at regular schools (and their parents) in China, yet its effect and cost-effectiveness are being hotly debated among researchers and policymakers. This study helps shed some light on this issue. Exploiting the panel structure of a large-scale survey that covers 28 counties/districts of China, this article applies a DID method (with propensity score matching in some analyses) to estimate the causal effects of private tutoring on students' learning outcomes.

Our estimation yields three important findings. First, subject-specific tutoring has a statistically significant and positive effect on students' performance on Chinese and mathematics tests, although these effects are at best modest in magnitude. Second, echoing findings of many recent qualitative studies (Liu, 2018; Zhang, 2014; Zhang & Bray, 2016), our analyses suggest that private tutoring improves students' academic performance mainly through enhancing their test-taking skills or through helping them accumulate more subject-specific knowledge, rather than through improving their general cognitive skills. This implies that to really serve the function of enhancing China's human capital formation, the private tutoring sector may need to be transformed so that it not only helps students understand better materials covered in regular classes, but also helps them develop more general cognitive skills. Finally, the effect of private tutoring is heterogenous across subsamples with different characteristics: it is larger for girls, low-performing students, and students with better-educated and wealthier parents. These heterogenous effects add a layer of complexity to the issue of educational inequality—while the remedial effect of private tutoring serves to reduce educational inequality, the fact that students from more advantageous backgrounds benefit more from private tutoring tends to increase educational inequality. More research on the issue of educational inequality associated with private tutoring is therefore needed to better inform policy.

In closing, a note on the limitations of this study is in order. First, due to data limitations, we are unable to test the key identification assumption underlying the DID method, the so-called 'parallel trend' assumption (i.e., the control and the treatment groups share a common trend of the outcome variable in the absence of treatment). As a second-best alternative, we included a large set of baseline covariates at various levels in estimation and resorted to propensity score matching in some analyses to help strengthen this assumption. Reassuringly, the robustness of estimation results across different estimators and specifications lend empirical support to the validity of our findings.

Second, although we have explored a number of potential channels through which private tutoring improves students' test scores, we are unable to pin down the exact ultimate channels (i.e., through improving students' test-taking skills versus the accumulation in subject-specific knowledge) due to the lack of direct information on these channels. Future studies that employ more detailed data on these channels are expected to be fruitful. Yet, in any case, our analysis suggests that private tutoring does not significantly improve students' general cognitive skills, even if it does raise students' test scores, in China.

Finally, our analysis focuses on only 8th graders (who were 7th graders at the baseline) in China. Yet the effects of private tutoring may well vary across grades—in particular, the effects of test-oriented private tutoring may be much larger for 9th graders, who are preparing for the high school entrance exam. Thus, future research involving students attending different grades may be fruitful in depicting a fuller picture of the effectiveness of private tutoring in China.

ACKNOWLEDGEMENTS

The authors thank Michael Delgado, Qiran Zhao, Chen Zhu for their helpful comments on an earlier version of this paper. We are also grateful for the constructive comments provided by the Associate Editor of the Journal, Professor John Gibson and an anonymous reviewer for revising the paper. All remaining errors are solely ours. This study received financial support from the Humanities and Social Sciences Fund of Ministry of Education of China [grant number: 18YJC790010]. Financial and logistic support from of the Center for Food and Health Economic Research (C'FHER) and the Beijing Food Safety Policy & Strategy Research Base at the China Agricultural University are also gratefully acknowledged.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Chinese National Survey Data Archive at http://www.cnsda.org/index.php?r=projects/view&id=72810330 and http://www.cnsda.org/index.php?r=projects/view&id=61662993.

DATA CITATION

National Survey Research Center; 2013; China Education Panel Survey; Chinese National Survey Data Archive; https://ceps.ruc.edu.cn/.

ORCID

Qihui Chen https://orcid.org/0000-0002-0944-0470

REFERENCES

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, *122*(3), 1235–1264. https://doi.org/10.1162/qjec.122.3.1235

Bhorkar, S., & Bray, M. (2018). The expansion and roles of private tutoring in India: From supplementation to supplantation. *International Journal of Educational Development*, *62*, 148–156. https://doi.org/10.1016/j.ijedudev.2018.03.003

Bray, M. (2013). Benefits and tensions of shadow education: Comparative perspectives on the roles and impact of private supplementary tutoring in the lives of Hong Kong students. *Journal of International and Comparative Education*, *2*(1), 18–30. https://doi.org/10.14425/00.45.72

Bray, M., & Kwok, P. (2003). Demand for private supplementary tutoring: Conceptual considerations, and socio-economic patterns in Hong Kong. *Economics of Education Review*, *22*(6), 611–620. https://doi.org/10.1016/S0272-7757(03)00032-3

Bray, M., Zhan, S., Lykins, C., Wang, D., & Kwo, O. (2014). Differentiated demand for private supplementary tutoring: Patterns and implications in Hong Kong secondary education. *Economics of Education Review*, *38*, 24–37. https://doi.org/10.1016/j.econedurev.2013.10.002

Buchmann, C. (2002). Getting ahead in Kenya: Social capital, shadow education, and achievement. *Research in the Sociology of Education*, *13*, 133–159. https://doi.org/10.1108/S1479-3539(2002)0000013008

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces*, *89*(2), 435–461. https://doi.org/10.1353/sof.2010.0105

Chen, Q. (2017). Impacts of late school entry on children's cognitive development in rural northwestern China: Does preprimary education matter? *Asia & the Pacific Policy Studies*, *4*(3), 586–601. https://doi.org/10.1002/app5.192

Chen, Q., Wang, X., & Zhao, Q. (2019). Appearance discrimination in grading? Evidence from migrant schools in China. *Economics Letters*, *181*, 116–119. https://doi.org/10.1016/j.econlet.2019.04.024

Chen, Y., & Feng, S. (2013). Access to public schools and the education of migrant children in China. *China Economic Review*, *26*, 75–88. https://doi.org/10.1016/j.chieco.2013.04.007

Cheo, R., & Quah, E. (2005). Mothers, maids and tutors: An empirical evaluation of their effect on children's academic grades in Singapore. *Education Economics*, *13*(3), 269–285. https://doi.org/10.1080/09645290500073746

Chinese Society of Education. (2016). Investigation report on current situations of after-school tutoring institutions and tutors in China. [In Chinese]. Retrieved from https://mp.weixin.qq.com/s/w84dYy77VoFznoOLxWUgrw?

Choi, J., & Cho, R. M. (2016). Evaluating the effects of governmental regulations on South Korean private cram schools. *Asia Pacific Journal of Education*, *36*(4), 599–621. https://doi.org/10.1080/02188791.2015.1064356

Damayanthi, B. W. R. (2018). Academic performance of students: The effect of class room level factors and private tutoring. *American Journal of Educational Research*, *6*(3), 262–269. https://doi.org/10.12691/education-6-3-13

Dang, H.-A. (2007). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review*, *26*(6), 683–698. https://doi.org/10.1016/j.econedurev.2007.10.003

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, *98*(1), 198–208. https://doi.org/10.1037/0022-0663.98.1.198

Ha, T. T., & Harpham, T. (2005). Primary education in Vietnam: Extra classes and outcomes. *International Education Journal*, *6*(5), 626–634.

Ha, Y., & Park, H. J. (2017). Can after-school programs and private tutoring help improve students' achievement? Revisiting the effects in Korean secondary schools. *Asia Pacific Education Review*, *18*(1), 65–79. https://doi.org/10.1007/s12564-016-9451-8

Huang, X., & Wei, Y. (2018). The private supplementary tutoring system in China: Findings from a national survey. In R. Wang (Ed.), *Annual report on new types of education suppliers (2017), basic education*. Beijing: Social Sciences Academic Press. [In Chinese]

Kim, K.-M., & Park, D. (2012). Impacts of urban economic factors on private tutoring industry. *Asia Pacific Education Review*, *13*(2), 273–280. https://doi.org/10.1007/s12564-011-9192-7

Kuan, P.-Y. (2011). Effects of cram schooling on mathematics performance: Evidence from junior high students in Taiwan. *Comparative Education Review*, *55*(3), 342–368. https://doi.org/10.1086/659142

Lai, F. (2010). Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. *Economics of Education Review*, *29*(3), 383–399. https://doi.org/10.1016/j.econedurev.2009.07.009

Liu, J. (2018). Review of Regulatory Policies on Private Supplementary Tutoring in China. *ECNU Review of Education*, *1*(3), 143–153. https://doi.org/10.30926/ecnuroe2018010307

Loyalka, P., & Zakharov, A. (2016). Does shadow education help students prepare for college? Evidence from Russia. *International Journal of Educational Development*, *49*, 22–30. https://doi.org/10.1016/j.ijedudev.2016.01.008

National Bureau of Statistics of China. (2018). *China statistical yearbook*. Beijing: China Statistics Press. [In Chinese]

Rubin, D. B. (1974). Estimating the causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Ryu, D., & Kang, C. (2013). Do private tutoring expenditures raise academic performance? Evidence from middle school students in South Korea. *Asian Economic Journal*, *27*(1), 59–83. https://doi.org/10.1111/asej.12002

Sato, Y. (Ed.) (2012). *100 years of Juku and 50 years of Juku associations*. Tokyo: Private Tutoring Federation. [In Japanese]

Stevenson, D. L., & Baker, D. P. (1992). Shadow education and allocation in formal schooling: Transition to university in Japan. *American Journal of Sociology*, *97*(6), 1639–1657. https://doi.org/10.1086/229942

Suryadarma, D., Suryahadi, A., Sumarto, S., & Rogers, F. H. (2006). Improving student performance in public primary schools in developing countries: Evidence from Indonesia. *Education Economics*, *14*(4), 401–429. https://doi.org/10.1080/09645290600854110

Tan, P.-L., & Gibson, J. (2017). Does outsourcing household production harm future human capital? Evidence from Malaysia. *Singapore Economic Review*, *62*(5), 959–981. https://doi.org/10.1142/S0217590815500745

Tansel, A., & Bircan, F. (2006). Demand for education in Turkey: A Tobit analysis of private tutoring expenditures. *Economics of Education Review*, *25*(3), 303–313. https://doi.org/10.1016/j.econedurev.2005.02.003

Thongphat, N. (2012). A survey of Thai student performance in mathematics and English: Evaluating the effect of supplementary tutoring. *Procedia Economics and Finance*, *2*, 353–362. https://doi.org/10.1016/S2212-5671(12)00097-4

Wang, R. (Ed.) (2017). *Annual report on new types of education suppliers (2017), basic education*. Beijing: Social Sciences Academic Press. [In Chinese]

Wang, X., Luo, R., Zhang, L., & Rozelle, S. (2017). The education gap of China's migrant children and rural counterparts. *Journal of Development Studies*, *53*, 1865–1881. https://doi.org/10.1080/00220388.2016.1274395

Zhang, W. (2014). The demand for shadow education in China: Mainstream teachers and power relations. *Asia Pacific Journal of Education*, *34*(4), 436–454. https://doi.org/10.1080/02188791.2014.960798

Zhang, W. (2019). Regulating private tutoring in China: Uniform policies, diverse responses. *ECNU Review of Education*, *2*(1), 25–43. https://doi.org/10.1177/2096531119840868

Zhang, W., & Bray, M. (2016). Shadow education: The rise and implications of private supplementary tutoring. In S. Guo & Y. Guo (Eds.), *Spotlight on China: Changes in education under China's market economy* (pp. 85–99). Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6209-881-7_6

Zhang, W., & Bray, M. (2017). Micro-neoliberalism in China: Public-private interactions at the confluence of mainstream and shadow education. *Journal of Education Policy*, *32*(1), 63–81. https://doi.org/10.1080/02680939.2016.1219769

Zhang, Y. (2013). Does private tutoring improve students' National College Entrance Exam performance? A case study from Jinan, China. *Economics of Education Review*, *32*, 1–28. https://doi.org/10.1016/j.econedurev.2012.09.008

Zhang, Y., & Liu, J. (2016). The effectiveness of private tutoring in China with a focus on class-size. *International Journal of Educational Development*, *46*, 35–42. https://doi.org/10.1016/j.ijedudev.2015.11.006

AUTHOR BIOGRAPHIES

Ms. Yuhe Guo is a PhD candidate at the China Agricultural University. Her research focuses on poverty reduction, public finance and the economics of education.

Dr. Qihui Chen received his PhD in Applied Economics from the University of Minnesota and is an Associate Professor at the China Agricultural University. He is currently serving as an Associate Editor of China Agricultural Economic Review, and a consultant for the World Bank. His research focuses on education, health and welfare issues in developing countries.

Ms. Shengying Zhai is a PhD candidate at the China Agricultural University. Her research focuses on human capital and rural development in China.

Ms. Chunchen Pei is currently pursuing her Master's degree in industrial economics. Her research focuses on the economics of education and nutrition in China.

How to cite this article: Guo Y, Chen Q, Zhai S, Pei C. Does private tutoring improve student learning in China? Evidence from the China Education Panel Survey. *Asia Pac Policy Stud*. 2020;7:322–343. https://doi.org/10.1002/app5.310

APPENDIX A

TABLE A1 Results of propensity scores estimation

	(1) Chinese tutoring	(2) Maths tutoring
Estimator	Logit	Logit
Age (months)	−0.008 (0.009)	−0.009 (0.007)
Boy (=1 if yes)	0.129 (0.112)	-0.348^{***} (0.085)
Cognitive ability score (standardised)	0.139^{*} (0.075)	−0.048 (0.057)
Birth order	−0.048 (0.137)	−0.136 (0.106)
Number of siblings	0.081 (0.109)	0.077 (0.085)
Father's education (years)	0.030 (0.026)	0.029 (0.019)
Mother's education (years)	0.060^{**} (0.025)	0.018 (0.018)
Household registration (=1 if rural)	0.067 (0.140)	-0.183^{*} (0.105)
Income level (dummies)	Yes	Yes
Class size	−0.011 (0.024)	0.041^{**} (0.020)
Subject teacher's education (years)	-0.212^{*} (0.123)	−0.060 (0.081)
Subject teacher's teaching experience (years)	−0.006 (0.011)	0.002 (0.008)
School fixed effects	Yes	Yes
Constant	1.188 (2.919)	−1.758 (1.892)
N	4,821	4,663
Pseudo R^2	0.094	0.137

Note: The covariates used to predict the propensity to receive the treatment (i.e., subject-specific tutoring) include the personal, family, and class/school characteristics observed at the baseline (Table 1, panel B). Standard errors in parentheses.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

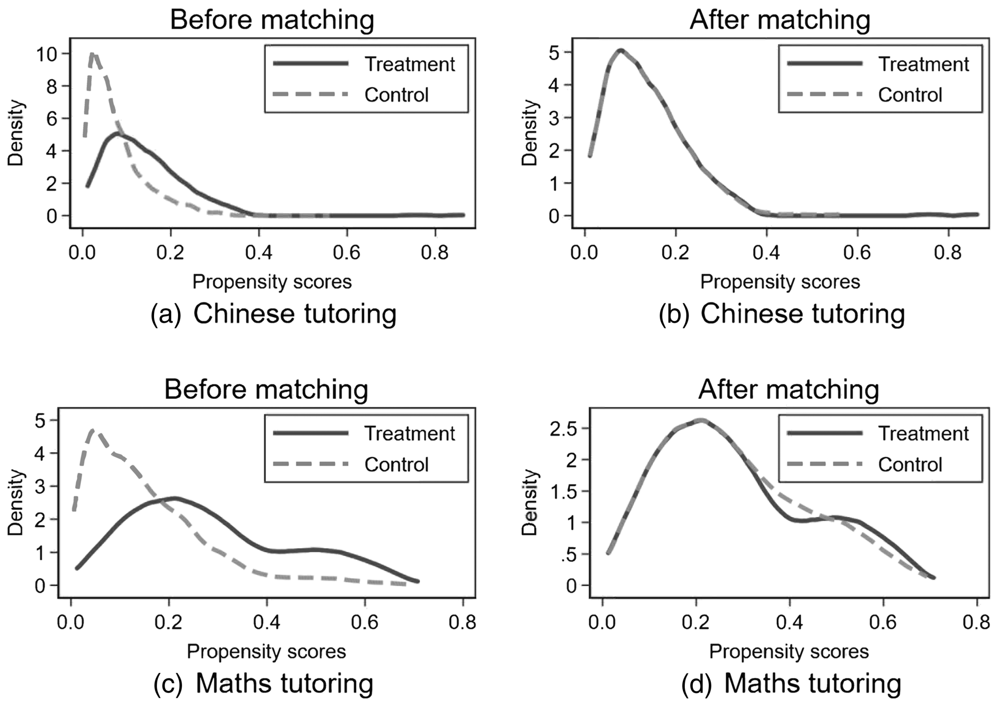

FIGURE A1 Estimated propensity scores