

课外补习对初中生学业成绩的影响效应

——基于 PSM-DID 方法的估计

管 振

[摘要] 学界对课外补习效应的研究还未形成一致结论。本文基于中国教育追踪调查(CEPS)基线和追访数据,采用 PSM-DID 方法,考察了初中阶段学生参与课外补习对学业成绩的影响及其在不同学生群体中的异质性。主要有以下发现:(1)父母教育水平、家庭经济水平和学业成绩显著提升学生参与数学和英语补习的概率;但家庭经济水平和学业成绩对学生参与语文补习的概率无显著影响。(2)语文、数学和英语补习都能够显著提升学生的学业成绩。其中,英语的补习效应最大,语文和数学的补习效应类似。(3)对于学业成绩优异的学生,各学科的补习效应都更大。因此,各学科的补习都会扩大学生间的成绩差距,不利于教育的均衡发展。

[关键词] 课外补习; 学业成绩; 教育公平; 双重差分

一、引言

课外补习是指为提高学生学业成绩而进行的、正规学校外的补充性教育活动(Bray and Kwok, 2003)。早在 20 世纪 60 年代,课外补习在亚、非、拉国家已普遍存在(Baker et al., 2001)。中国因社会发展转型,课外补习市场崛起较晚(李佳丽, 2018)。20 世纪 80 年代后期,一些城市家庭父母请老师或大学生到家中为孩子辅导,成为中国现代课外补习的起源。一直到现在,这种补习形式仍然存在,但已经被更广泛的小组和大班制的校外培训所淹没(陈全功, 2012)。近些年来,我国出台一系列政策鼓励减轻中小学生的校外学习负担,但课外补习率仍保持在较高水平(张冰和程天君, 2019)。2012 年全国有约 24.6% 的义务教育在校生参加了课外补习(薛海平, 2015), 2017 年的补习率更是超过 30%(魏易和薛海平, 2021)。

[收稿日期] 2021-10-25

[作者简介] 管振,北京师范大学经济与工商管理学院/首都教育经济研究院,电子邮箱地址: guanzhen0304@163.com。

课外补习不仅是家庭教育竞争的重要一环，也是教育机会不均等的重要体现，受到政府和学界的广泛关注(Dang and Rogers, 2008)。学界对课外补习的研究主要集中在两个方面，分别是参与课外补习的影响因素和其对学业成绩的影响。对于前者，大多数研究发现，家庭经济水平越优、父母的教育程度越高，子女参与课外补习的概率越大(Stevenson and Baker, 1992; 洪岩璧和赵延东, 2014; 薛海平和赵阳, 2020)。因此，课外补习也被认为具有社会阶层复制的作用(薛海平和李静, 2016)。根据学业成绩较优和较劣的学生有更大概率参与课外补习，可以将课外补习分为培优型和补差型(Baker et al., 2001)。研究发现，因为儒家社会，如韩国、日本和中国等对等级考试更加重视，以培优型为主(Entrich, 2014; Kim et al., 2010)。而欧美国家，如美国、英国、以色列等的补习活动多是依靠公共财政，对学习成绩不佳或贫困的学生提供支持，以补差型为主(Baker et al., 2001)。

诸多学者研究了语文、数学和英语补习对相应学科成绩的影响，还未形成统一的结论。有的学者研究发现各学科补习都有效(薛海平, 2016)。与此相反，有的学者研究发现各学科补习都无效(孙伦轩和唐晶晶, 2019)。也有学者研究发现部分学科补习有效，部分无效，如李波(2018)发现尽管数学和英语补习有效，但是语文补习无效；胡咏梅等(2021)发现语文和英语补习有效，但是数学补习无效。除了不同个人特征和家庭背景等学生的补习效应存在异质性之外，没有形成统一结论的主要原因是，有些研究中存在内生性问题(胡咏梅等, 2015)。

学界常常采用普通最小二乘(OLS)、倾向得分匹配(PSM)、两阶段最小二乘(2SLS)、双重差分(DID)和倾向得分匹配-双重差分(PSM-DID)等方法研究课外补习对学业成绩的影响(李波和黄斌, 2020)。OLS和PSM方法的基本思想相同：依据选取的变量平衡补习和未补习学生间的异质性差异，从而得出在其他可观测特征相同的情况下，课外补习对学业成绩的影响。然而，其估计的有效性易受到遗漏变量，特别是学生的个人能力和初始成绩的影响。如果学业成绩优异的学生更倾向于参与课外补习，遗漏初始成绩变量可能会高估补习效应；反之，低估补习效应。尽管部分学者在模型中加入了初始成绩变量，但是学生的个人能力仍然无法控制(薛海平等, 2014; 李佳丽和胡咏梅, 2017)。

为了克服遗漏变量问题，有些学者采用2SLS方法进行估计。其基本思想是：依据与核心解释变量(课外补习)相关，而与随机误差项不相关的外生变量，进行局部随机分组，进而得出课外补习对学业成绩的因果影响。常见的工具变量有公立学校收取的辅导费、家离补习班的距离、班级内学生的补

习比例等(Dang, 2007; Zhang, 2013; Suryadarma et al., 2006; 杨莉, 2020)。在相关性和外生性条件满足的情况下, 2SLS方法能够估计出课外补习的局部处理效应。然而, 工具变量的外生性常常难以满足; 而且由于其估计的是局部处理效应, 外部有效性也常常受到质疑。

近年来, 有些学者尝试采用DID方法估计课外补习对学业成绩的影响。其基本思想是: 如果补习的学生未补习, 他(她)的成绩将和未补习的学生有着相同的变化(平行趋势假设), 则补习学生成绩的变化减去未补习学生成绩的变化, 即为补习效应。为了使平行趋势假设得以满足, Heckman et al. (1997)提出使用PSM-DID方法进行估计。PSM-DID方法的基本思想是: 首先, 使用PSM方法估计进行样本筛选, 以此平衡补习和未补习学生间的可观测特征差异; 然后, 利用筛选后的样本再进行DID估计。部分学者使用此方法考察了课外补习对学业成绩的影响, 但是还存在一些不足; 其中最主要是未将基期样本限定在未参与补习的学生当中, 导致估计的结果为两期补习效应的差异(孙伦轩和唐晶晶, 2019; 庞晓鹏等, 2017)。

鉴于目前在补习效应的研究中还未形成一致的结论, 而且方法应用上还存在一些不足, 本文使用中国教育追踪调查(CEPS)基线和追访数据, 采用PSM-DID方法估计语文、数学和英语补习对相应学业成绩的影响; 并考察其不同学生群体间的异质性。本研究的理论意义在于: 完善了PSM-DID方法在补习效应研究中的应用, 进一步深化对补习效应的认识。现实意义在于: 研究结论既有助于学生和家长合理分配课外学习时间; 也有助于政府更有效地配置教育资源, 促进教育公平和健康发展。

本文接下来的安排如下, 第二部分介绍模型设定、数据来源和处理结果; 第三部分报告课外补习的影响因素以及补习效应的估计结果; 第四部分进一步分析补习效应的异质性以及放弃课外补习对学业成绩的影响; 第五部分总结研究结论, 并提出相应的政策建议。

二、模型与数据

(一) 补习效应的识别

本文使用七年级和八年级学生样本, 估计七、八年级间的语文、数学和英语补习对各学科成绩的影响。因为各学科补习效应的识别策略相同, 下面将做一般化的讨论。

DID方法识别补习效应的思路如图1所示。首先, 由于估计七、八年级间的补习效应, 我们考虑学生处于七、八年级两个时点的学业成绩。其次,

将样本限定在七年级前未参与补习的学生中，以使初始样本全部是未受处理的学生。再次，根据有部分学生在七、八年级间参与补习($treat=1$)，有部分未参与($treat=0$)，把学生分成两类。基于以上设定，在七、八年级间补习的学生的成绩变化为 C_0F_0 。因为即使未参与补习，学生的成绩在此期间也可能发生变化(假设为 E_0F_0)，所以补习效应为： $C_0E_0 = C_0F_0 - E_0F_0$ 。但是， E_0F_0 无法观测。为此，用七、八年级间未补习学生的成绩变化(即 D_0G_0)代替 E_0F_0 作为反事实。这就要求满足： $B_0D_0 // A_0E_0$ 以及 $D_0G_0 = E_0F_0$ (即平行趋势假设)。

为了使平行趋势假设得以满足，使用七年级的学生样本进行 PSM 估计，来筛选出其他特征相似的补习和未补习的学生；然后再进行 DID 估计。若平行趋势假设得以满足， $C_0E_0 = C_0F_0 - E_0F_0 = C_0F_0 - D_0G_0$ 就反映出七、八年级间的课外补习对学业成绩的因果影响。以上即为 PSM-DID 估计补习效应的基本思路。

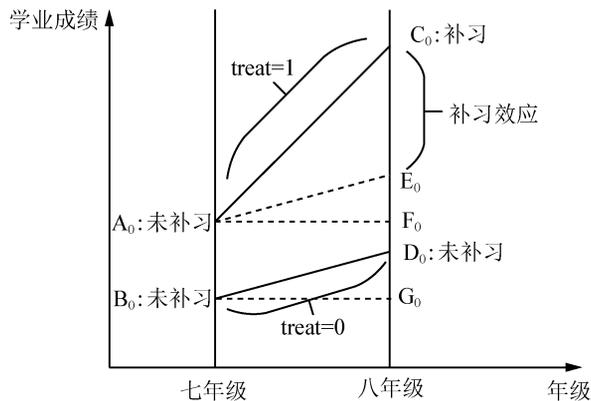


图 1 DID 方法对补习效应的识别

(二) 模型设定

1. DID 模型

根据 DID 方法对补习效应识别的思路，本文采用以下模型估计语文、数学和英语补习对相应学科成绩的影响：

$$score_{ijt} = \alpha_{0j} + \alpha_{1j}treat_{ij} + \alpha_{2j}after_{ijt} + \alpha_{3j}(treat_{ij} * after_{ijt}) + X_{ijt}\alpha_{4j} + \mu_{ijt} \quad (1)$$

其中，下标 j 分别表示语文、数学或英语学科。 $t=0$ 和 $t=1$ 分别表示学生处于七、八年级。在进行 j 学科的估计中，仅仅使用了七年级前未参加过 j 学科补习的样本。当学生 i 在七、八年级间参加学科 j 的课外补习时， $treat_{ij}=1$ ；反之， $treat_{ij}=0$ 。 $t=0$ 和 $t=1$ 分别对应于 $after_{ijt}=0$ 和 $after_{ijt}$

$=1$ 。 $score_{ijt}$ 表示学生相应的考试成绩。 X_{ijt} 表示模型中的控制变量集，包含学生的个人特征(性别、年龄、独生子女和城乡身份)、家庭背景(父母教育水平、家庭经济水平)、学校情况(公立学校、学校质量)、学生的寄宿状况和区县固定效应等。 μ_{ijt} 表示对学生成绩有影响的随机扰动项。

模型(1)与图 1 的内容一一对应。 $treat=0, after=0$ 意味着七、八年级间未补习的学生在七年级的状态，与图 1 中的 B_0 点相对应；此时， $score_1=r$ ；其中， $r=\alpha_{0j}+X_{ijt}\alpha_{4j}+\mu_{ijt}$ 。^① 同理， $treat=0, after=1$ 与图 1 中的 D_0 点相对应；此时， $score_2=\alpha_2+r$ 。 $treat=1, after=0$ ，与图 1 中的 A_0 点相对应；此时， $score_3=\alpha_1+r$ 。 $treat=1, after=1$ ，与图 1 中的 C_0 点相对应；此时， $score_4=\alpha_1+\alpha_2+\alpha_3+r$ 。各状态学生对应的学业成绩如表 1 所示。

C_0F_0 反映了 C_0 与 A_0 的成绩差，即：

$$C_0F_0=score_4-score_3=(\alpha_1+\alpha_2+\alpha_3+r)-(\alpha_1+r)=\alpha_2+\alpha_3$$

D_0G_0 反映了 D_0 与 B_0 的成绩差，即：

$$D_0G_0=score_2-score_1=(\alpha_2+r)-r=\alpha_2$$

根据前文分析，补习效应 C_0E_0 为：

$$C_0E_0=C_0F_0-D_0G_0=(\alpha_2+\alpha_3)-\alpha_2=\alpha_3$$

即模型(1)中交互项的系数估计值反映了各学科补习对学业成绩的影响。

表 1 DID 模型中各状态学生的学业成绩

	八年级(after=1)成绩	七年级(after=0)成绩	difference
七、八年级间补习($treat=1$)	$\alpha_1+\alpha_2+\alpha_3+r$	α_1+r	$\alpha_2+\alpha_3$
七、八年级间未补习($treat=0$)	α_2+r	r	α_2
<i>difference</i>	$\alpha_1+\alpha_3$	α_1	α_3

2. PSM-DID 模型

补习效应的正确识别，依赖于平行趋势假设得以满足。为此，首先使用七年级的横截面数据进行 PSM 估计来筛选样本，以平衡补习和未补习学生间的特征差异；然后再利用模型(1)进行 DID 估计。PSM 估计的基本模型为：

$$P(treat_{ij}=1 | X_{ij0}, score_{ij0}) = \frac{\exp(\beta_{0j} + X_{ij0}\beta_{2j} + score_{ij0}\beta_{2j} + \epsilon_{ij0})}{1 + \exp(\beta_{0j} + X_{ij0}\beta_{2j} + score_{ij0}\beta_{2j} + \epsilon_{ij0})} \quad (2)$$

其中，下标 $t=0$ ，意味着仅包含七年级的样本。跟前文一样，在进行 j 学科的估计中，仅仅使用了七年级前未参加过 j 学科补习的样本。其他变量的定义与模型(1)相同。此外，为了平衡七、八年级间补习和未补习学生的初

^① 为了表达简洁，在此处的讨论中省略了部分下标；下同。

始成绩差异,在模型中也加入了学生七年级的成绩 $score_{ij0}$ 。 ϵ_{ij0} 为随机误差项。

通过估计模型(2),可以得出所有样本的倾向得分值。然后依据不同的匹配策略(如k近邻匹配、半径匹配、核匹配、马氏匹配等)进行样本筛选后,再采用模型(1)进行DID估计。即为本文所采用的PSM-DID估计模型。此外,模型(2)也被用于分析学生参与语文、数学和英语补习的影响因素。

(三)数据来源、处理与描述性统计

1. 数据来源

本文使用的数据来源源于中国教育追踪调查(CEPS)的2013—2014学年基线和2014—2015学年追访数据。该数据是由中国人民大学中国调查与数据中心设计与实施的、具有全国代表性的大型追踪调查项目。其从全国随机抽取了28个县级单位(县、区、市)作为调查点,总共调查了112所学校、438个班级。在基线调查中,在七年级总共调查了10279名学生;一年后,成功追访9449名,追访率为91.9%。我们把基线调查中七年级的学生样本和一年后的追访样本(学生升为八年级)进行匹配与合并,即为本文使用的全部样本来源。

2. 数据处理

根据模型设定,选择的变量包括学生的基本信息、家庭背景、学校情况、语文、数学及英语的补习状况和考试成绩等。各变量的定义如表2所示。需要说明的是,在七年级和八年级分别询问了学生“参加了哪些兴趣班或课外辅导班”和“最近一年来,参加了哪些兴趣班或课外辅导班”,答题选项包含“没参加、数学、语文、英语及其他与课业无关的兴趣培训等”。本文据此识别出学生在七年级前和以及七、八年级间是否参加过某学科的课外补习。在各学科考试成绩的处理上,参考孙伦轩和唐晶晶(2019)的研究,标准成绩按照学校进行标准化为:(学生成绩-最低分)/(最高分-最低分)。此外,父母教育水平由问卷调查的父母最高教育程度转化而来,各阶段教育年限分别为:小学6年,初高中各3年,大专3年,本科4年,研究生3年,博士4年。

此外,为了消除特差生和特优生的影响,仅保留了在七年级语文、数学和英语成绩为1%分位到99%分位之间的学生。

3. 描述性统计

在删除了各变量的缺失值和异常值后,共得到12474个样本。其中,七年级前未参与语文补习的样本最多,约10960个;而未参与数学和英语补习的人数相当,分别为9584和9286个。各变量描述性统计结果如表2所示。可以发现,在七年级前未参与数学补习的学生中,约17.8%在七、八年级间

参与了数学补习；相应的英语补习比例约为 13.2%；而仅有约 7.5% 的学生转向了语文补习。从以上可以看出，初中学生更加青睐于数学和英语补习，而对语文补习有所忽视。在学生的个人特征上，学生平均年龄约为 12.5 岁，男、女生各占约 50%，其中约 45% 的学生为独生子女。在家庭背景上，父母最高教育年限平均约为 11 年，经济水平约为 3，即一般水平。在学校特征上，大多数学校为公立学校，占比约 94%；学校质量评价约为 4，处于良好的水平。

表 2 变量定义与描述性统计

	变量名	定义	样本量	均值	标准差
被解释变量	语文成绩	按学校进行标准化的语文成绩	12474	0.614	0.216
	数学成绩	按学校进行标准化的数学成绩	12474	0.633	0.258
	英语成绩	按学校进行标准化的英语成绩	12474	0.623	0.264
核心解释变量	语文补习	学生七、八年级间参与了语文补习为 1，否则为 0	10960	0.075	0.263
	数学补习	学生七、八年级间参与了数学补习为 1，否则为 0	9584	0.178	0.383
	英语补习	学生七、八年级间参与了英语补习为 1，否则为 0	9286	0.132	0.339
	男生	男生为 1，否则为 0	12474	0.504	0.500
	年龄	调查年份 - 出生年份	12474	12.516	0.692
	独生子女	独生子女为 1，否则为 0	12474	0.451	0.498
	父母教育水平	父母的最高教育年限	12474	11.037	3.002
控制变量	家庭经济水平	家庭的经济状况；1-5 分别表示从非常困难到很富裕	12474	2.985	0.583
	城市居民	城镇户口为 1，否则为 0	12474	0.479	0.500
	寄宿	学生在周一到周四在学校寄宿为 1，否则为 0	12474	0.318	0.466
	公立学校	学生所在学校是公立学校为 1，否则为 0	12474	0.941	0.236
	学校质量	学生的学校在县里的排名；1-5 分别表示最差到最好	12474	3.974	0.829

三、基本回归结果

(一) 参与课外补习的影响因素

表3报告了学生参与语文、数学和英语补习的影响因素回归结果。系数已转化为平均边际效应。总的来看,父母教育水平、家庭经济水平和学业成绩对学生参与数学和英语补习都有着显著的正向影响。这体现出课外补习机会分配的不均衡。如果补习效应为正,则数学和英语补习可能造成学生学业成绩进一步分化。

从父母教育水平来看,其对子女选择语文、数学和英语补习都有着显著的正向影响。从家庭经济水平来看,尽管其对学生的数学和英语补习有着显著的正向影响,但是对语文补习的影响却不显著。说明经济条件优越的家庭更加重视孩子的英语和数学补习,但各经济阶层对语文补习的重视程度没有较大差异。从学业成绩来看,数学和英语成绩分别对学生选择数学和英语补习有着显著的正向影响,但是语文成绩对选择语文补习却无显著影响。这说明我国的数学和英语补习主要是“培优型”;但是语文补习未体现明显的“培优”或“补差”特征。

表3 参与补习的影响因素分析结果

变量名	(1)	(2)	(3)
	语文补习	数学补习	英语补习
男生	0.008 (0.008)	-0.035*** (0.011)	0.001 (0.010)
年龄	-0.010* (0.006)	-0.018** (0.009)	-0.017** (0.008)
独生子女	-0.002 (0.008)	0.008 (0.013)	0.015 (0.011)
父母教育水平	0.007*** (0.002)	0.007*** (0.002)	0.007*** (0.002)
家庭经济水平	0.001 (0.007)	0.032*** (0.011)	0.025*** (0.010)
城市居民	0.011 (0.009)	0.019 (0.013)	-0.005 (0.012)
寄宿	-0.011 (0.013)	-0.039** (0.018)	-0.002 (0.016)
公立学校	-0.015 (0.017)	-0.000 (0.028)	0.053** (0.027)

续表			
变量名	(1)	(2)	(3)
	语文补习	数学补习	英语补习
学校质量	0.013** (0.006)	0.013 (0.008)	0.014** (0.007)
语文成绩	-0.001 (0.019)		
数学成绩		0.042* (0.024)	
英语成绩			0.047** (0.023)
区县固定效应	控制	控制	控制
样本量	5313	4792	4643
<i>Pseudo R</i> ²	0.067	0.117	0.100

注：括号内报告了稳健标准误；*、**、*** 分别表示在 10%、5% 和 1% 的水平上显著。

(二) 倾向得分匹配效果检验

1. 平衡性检验

在进行 DID 估计之前，首先使用了倾向得分匹配进行样本筛选。倾向得分匹配效果可以从协变量的平衡性及样本的共同支撑区域两个方面来考察。表 4 报告了基于 1:3 近邻匹配后的平衡性检验结果。可以发现，所有协变量在匹配之后，处理组和对照组的标准化偏差都小于 10%；t 检验在 0.10 的水平上无法拒绝不存在组间差异的原假设；协变量联合显著性检验在 0.10 的水平上无法拒绝所有协变量的系数都为 0 的原假设。这说明倾向得分匹配很好的平衡了补习和未补习学生间的可观测性特征差异。

2. 共同支撑检验

图 2 分别报告了涉及语文、数学和英语补习的倾向得分匹配的倾向值的共同支撑区间。可以发现，在所有的倾向得分匹配中，只有极少的样本在共同取值范围外，因此倾向得分匹配仅损失少量样本。总的来看，倾向得分匹配取得较好的效果。

表 4 倾向得分匹配效果检验结果

变量名	匹配状态	(1)	(2)	(3)	(4)	(5)	(6)
		语文补习		数学补习		英语补习	
		偏差 (%)	t 检验 (t 值)	偏差 (%)	t 检验 (t 值)	偏差 (%)	t 检验 (t 值)
男生	匹配前	4.0	0.78	-12.7	-3.38***	-4.3	-0.98
	匹配后	5.0	0.72	1.8	0.38	-0.8	-0.15

续表

变量名	匹配状态	(1)	(2)	(3)	(4)	(5)	(6)
		语文补习		数学补习		英语补习	
		偏差 (%)	t 检验 (t 值)	偏差 (%)	t 检验 (t 值)	偏差 (%)	t 检验 (t 值)
年龄	匹配前	-18.3	-3.34***	-28.8	-7.20***	-26.6	-5.68***
	匹配后	0.3	0.04	-0.2	-0.04	1.3	0.25
独生子女	匹配前	25.3	4.92***	37.4	10.00***	34.6	8.08***
	匹配后	-1.4	-0.19	-2.2	-0.45	3.4	0.58
父母教育水平	匹配前	41.2	8.46***	48.1	13.06***	38.6	9.38***
	匹配后	-0.9	-0.13	4.6	0.92	0.4	0.06
家庭经济水平	匹配前	17.8	3.48***	34.3	8.84***	32.1	7.18
	匹配后	-1.9	-0.27	0.5	0.10	4.3	0.79
城市居民	匹配前	-1.9	5.51***	40.7	10.79***	25.5	5.93
	匹配后	2.8	0.41	3.7	0.77	2.1	0.36
寄宿	匹配前	-12.4	-2.36**	-40.2	-10.12***	-20.6	-4.64***
	匹配后	2.1	0.31	2.2	0.51	-0.0	0.01
公立学校	匹配前	-2.3	-0.45	15.0	3.62***	16.7	3.45***
	匹配后	2.1	0.29	0.5	0.13	0.2	0.04
学校质量	匹配前	25.1	4.62	24.9	6.37***	22.5	5.01***
	匹配后	-1.0	-0.15	1.7	0.37	6.1	1.12
语文成绩	匹配前	0.5	0.10				
	匹配后	-3.2	-0.45				
数学成绩	匹配前			19.3	4.86***		
	匹配后			3.6	0.76		
英语成绩	匹配前					8.4	1.85*
	匹配后					1.5	0.26
LR chi2	匹配前	86.40***		299.87***		156.93***	
	匹配后	1.62		3.09		2.33	

注：*、**、*** 分别表示在 10%、5% 和 1% 的水平上显著。

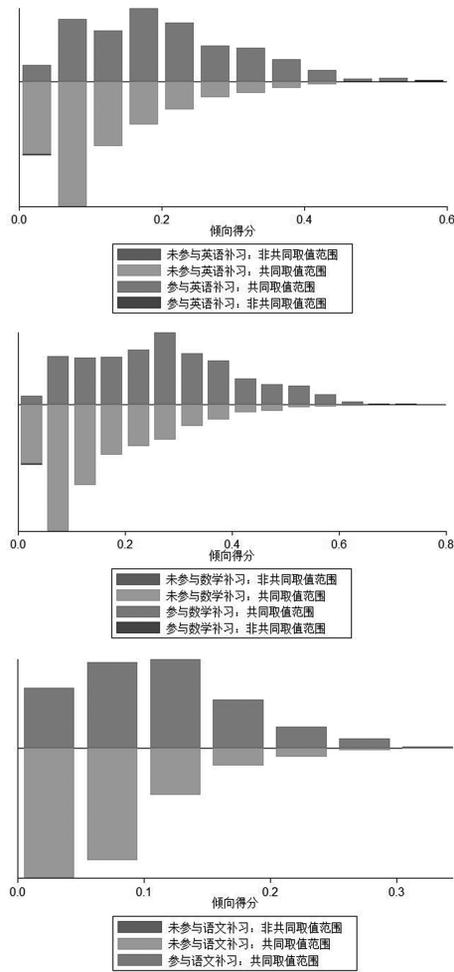


图 2 倾向得分的共同取值范围

(三) 课外补习对学业成绩影响的估计结果

表 5 分别报告了语文、数学和英语补习对相应学科成绩影响的 DID 和 PSM-DID 估计结果。总的来看, 无论对于语文、数学和英语补习, 两种方法的估计结果都类似, 而且对相应的学业成绩都有着显著的正向影响。这说明语文、数学和英语补习都有效, 与薛海平(2016)的研究结论一致。分别看来, 英语补习对英语成绩的提升作用最大, 使标准化成绩提升约 0.05 分。而语文和数学补习效应类似, 都使相应标准化成绩提升约 0.02 分。

但是部分学者使用 PSM-DID 方法估计得出各学科补习都无效的结论(孙伦轩和唐晶晶, 2019; 庞晓鹏等, 2017)。这可能因为他们未将基期样本限定在未补习的学生当中。

续表

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	语文			数学			英语		
	半径匹配	核匹配	马氏匹配	半径匹配	核匹配	马氏匹配	半径匹配	核匹配	马氏匹配
样本量	2790	10626	2696	4776	9580	5178	3768	9282	3596
R^2	0.158	0.126	0.131	0.146	0.154	0.157	0.236	0.267	0.231

注：括号内报告了稳健标准误；*、**、*** 分别表示在 10%、5% 和 1% 的水平上显著；模型所加入的控制变量与表(5)相同。

四、进一步的分析

(一)异质性分析

这里分别考察了各学科补习效应的性别、城乡、父母教育水平、家庭经济水平和原始成绩差异，表 7 报告了相应的估计结果。总的来看，在学生的原始成绩方面，无论对于语文、数学还是英语学科，学业成绩优异的学生的补习效应要高于学业成绩较差的学生，一定程度上说明各学科补习都会扩大学生间的成绩差异。而在性别、城乡、父母教育水平和家庭经济水平方面，并未出现某一学生群体在所有的学科补习中表现出完全的优势或劣势。例如，尽管父母教育水平高的学生的数学和英语补习效应更小，但是语文补习效应却更大。尽管家庭经济条件优越的学生的语文和数学补习效应更大，但是英语补习效应却更小。

从性别来看，语文、数学和英语补习对女生都有效；而对男生来说，尽管英语补习有效，但是语文和数学补习却无效。从学生的城乡身份来看，对于农村学生，数学和英语补习有效，但是语文补习却无效；而对于城镇学生，数学补习无效，语文和英语补习有着略微的效果。从父母教育水平来看，对于父母教育程度为高中以下的学生，数学和英语补习有效，而语文补习无效；而父母教育程度为高中及以上的学生，语文和英语补习有效，但是数学补习无效。从家庭经济水平来看，对家庭经济条件较差的学生，仅仅英语补习有效；但是家庭经济条件较优的学生，各学科的补习都有效。从原始分数来看，对于成绩较差的学生，仅仅英语补习有效；但是对于成绩较好的学生，各学科的补习都有效。

表7 异质性分析结果

		语文		数学		英语	
		DID	PSM-DID	DID	PSM-DID	DID	PSM-DID
性别	女生	0.026 [*] (0.015)	0.033 [*] (0.017)	0.027 ^{**} (0.012)	0.032 ^{**} (0.014)	0.082 ^{***} (0.014)	0.048 ^{***} (0.016)
	男生	0.028 [*] (0.016)	0.014 (0.018)	0.012 (0.012)	0.013 (0.014)	0.054 ^{***} (0.014)	0.051 ^{***} (0.016)
城乡身份	农村	0.029 (0.018)	0.021 (0.020)	0.061 ^{***} (0.014)	0.058 ^{***} (0.016)	0.082 ^{***} (0.015)	0.078 ^{***} (0.017)
	城市	0.035 ^{**} (0.015)	0.027 [*] (0.016)	0.003 (0.012)	0.001 (0.014)	0.051 ^{***} (0.015)	0.026 (0.017)
父母教育水平	高中以下	0.026 (0.017)	0.003 (0.018)	0.053 ^{***} (0.013)	0.044 ^{***} (0.015)	0.085 ^{***} (0.016)	0.069 ^{***} (0.018)
	高中及以上	0.040 ^{***} (0.015)	0.040 ^{**} (0.016)	0.004 (0.011)	0.012 (0.013)	0.050 ^{***} (0.013)	0.037 ^{**} (0.015)
家庭经济水平	一般以下	0.001 (0.032)	-0.011 (0.035)	-0.013 (0.042)	-0.000 (0.049)	0.091 ^{**} (0.040)	0.088 [*] (0.051)
	一般及以上	0.028 ^{**} (0.012)	0.028 ^{**} (0.013)	0.024 ^{***} (0.009)	0.023 ^{**} (0.011)	0.061 ^{***} (0.011)	0.046 ^{***} (0.012)
原始成绩	中位数以下	0.007 (0.015)	0.009 (0.017)	0.002 (0.013)	0.013 (0.015)	0.047 ^{***} (0.014)	0.046 ^{***} (0.016)
	中位数以上	0.046 ^{***} (0.016)	0.038 ^{**} (0.017)	0.043 ^{***} (0.012)	0.031 ^{**} (0.013)	0.093 ^{***} (0.015)	0.059 ^{***} (0.016)

注：括号内报告了稳健标准误；*、**、*** 分别表示在 10%、5% 和 1% 的水平上显著；模型中所加入的控制变量与表(5)相同；参数反映了各学科的(补习 * 八年级)交互项的系数估计值。

(二)放弃补习对学业成绩的影响

前文讨论了参与补习对学生学业成绩产生的影响。作为对照，这里考察放弃补习应该对学业成绩的影响。首先，在研究放弃某学科补习对学业成绩的影响中，将样本限定在七年级前参与此学科补习的样本中。其次，学生在七、八年级间不补习(即放弃补习)定义为 1，否则为 0。估计方法与前文相同，采用 PSM-DID 估计。

表 8 分别报告了相关的估计结果。总的来看，无论对于语文、数学还是英语学科，交互项的系数估计值都为负，说明放弃某学科补习对此学科的成绩

绩有着负向影响，与前文估计的参与补习对学业成绩影响的方向相反。具体来看，放弃语文补习对语文成绩的负向影响较小，而且在 0.10 的水平上不显著，说明放弃语文，短期内不会降低语文成绩。而放弃数学和英语补习对相应成绩的负向影响较大，且至少都在在 0.10 的水平上显著。因此放弃数学和英语补习短期内会较大的降低其数学和英语成绩。

表 8 放弃课外补习对学业成绩影响的估计结果

	(1)	(2)	(3)	(4)	(5)	(6)
	语文		数学		英语	
	DID	PSM-DID	DID	PSM-DID	DID	PSM-DID
放弃补习	-0.041** (0.016)	-0.037** (0.016)	-0.029** (0.012)	-0.012 (0.013)	-0.013 (0.011)	-0.012 (0.011)
八年级	0.031** (0.013)	0.032** (0.014)	-0.009 (0.008)	0.001 (0.010)	-0.095*** (0.008)	-0.098*** (0.009)
八年级 * 放弃补习	-0.019 (0.017)	-0.020 (0.018)	-0.031** (0.013)	-0.044*** (0.015)	-0.029** (0.011)	-0.022* (0.012)
控制变量	控制	控制	控制	控制	控制	控制
样本量	1514	1421	2890	2098	3188	2780
R ²	0.159	0.162	0.112	0.118	0.237	0.232

注：括号内报告了稳健标准误；*、**、*** 分别表示在 10%、5% 和 1% 的水平上显著；模型中所加入的控制变量与表(5)相同。

五、结论和政策建议

课外补习不仅是教育竞争的重要一环，也是教育机会不均等的重要表现。对课外补习的研究具有重要的理论和现实意义。本文基于中国教育追踪调查(CEPS)2013—2014 学年基线和 2014—2015 学年追访调查数据，采用 PSM-DID 方法分别考察了初中生的语文、数学和英语补习对其相应学业成绩的影响。主要有以下发现：

从参与课外补习的影响因素来看，父母教育水平和家庭经济水平对学生参与数学和英语补习有着显著的正向影响。而尽管父母教育水平对参与子女语文补习有着正向影响，但是家庭经济水平对其参与语文补习的影响却不显著。由于家庭背景越好的学生参与数学和英语补习概率越大，说明数学和英语补习加剧了学生获得义务教育资源的不均衡，与我国政府推行的义务教育公平政策相悖。此外，我国的数学和英语补习是培优型的；但是语文补习未

呈现显著的培优或补差特征。

从各学科的补习效应来看,无论对于语文、数学还是英语补习都能够显著提升相应学科的学业成绩。而且,英语补习对学生的英语成绩提升作用最大,而语文和数学补习对相应学科成绩的提升作用类似。这一结论在采用不同的倾向得分匹配方法下依然稳健。作为对照,放弃某学科的补习对此学科的学业成绩有着负向影响。

从补习效应的异质性来看,某一家庭背景的学生并未在所有的学科补习中表现出完全的优势或劣势。例如,尽管父母教育水平高的学生的数学和英语补习效应更小,但是语文补习效应却更大。尽管家庭经济条件优越的学生的语文和数学补习效应更大,但是英语补习效应却更小。此外,学业成绩优异的学生的补习效应要高于学业成绩较差的学生,这一定程度上说明各学科补习都会扩大学生间的成绩差异。

根据以上研究结论,本文提出以下两点政策建议:

其一,政府应大力支持学校开展课后服务,切实减弱学生的校外补习需求。校外补习兴起的原因很大一部分来自于学生在校时间的减少。在中国的“中考”和“高考”等教育制度下,学生和家长会积极利用教育资源以提升学生的学业成绩。于是,校外补习成为了多数人的选择。但是本文的研究发现,总的来看,学业成绩优异的学生更倾向于参与校外补习,而且补习效应更大。这势必造成扩大学生间的成绩差异,不利于教育的均衡化发展。因此,政府应鼓励学校开展课后服务,从而减弱学生的校外补习需求。

其二,政府应对弱势群体学生提供辅导资助,促进教育的公平和健康发展。课外补习具有一定的社会复制功能,反映了经济社会差距和社会不平等状况,是城乡差距和贫富差距的具体映像(陈全功,2012)。本文的研究结果也发现,总的来看,家庭背景更优的学生更容易选择补习。由于各学科的补习效应都为正,因此课外补习活动会扩大不同阶层学生的成绩差距。为此,政府应该借鉴西方国家补习经验,把补习活动看作对公立教育的有益补充,向弱势群体学生提供一定的补习资助。

当然,本文的研究还存在一定的不足。尽管本文在 DID 估计之前,采用 PSM 方法平衡补习和未补习学生的可观测特征差异,以使得平行趋势假设得以满足。但是由于无法获取更多年份的追踪数据,不能对平行趋势假设进行检验。这也可以作为进一步完善研究的方向。

[参考文献]

- 陈全功,2012:《补习教育的地域延展及其社会效应分析》,《比较教育研究》第3期。
洪岩璧、赵延东,2014:《从资本到惯习:中国城市家庭教育模式的阶层分化》,《社会学

研究》第 4 期。

胡咏梅、范文凤、丁维莉, 2015:《影子教育是否扩大教育结果的不均等——基于 PISA 2012 上海数据的经验研究》,《北京大学教育评论》第 3 期。

胡咏梅、元静、王亚男, 2021:《课外补习对初中生不同学科学业成绩的异质性影响》,《首都师范大学学报(社会科学版)》第 1 期。

李波、黄斌, 2020:《破解教育生产“黑箱”:教育生产函数研究的评述与展望》,《华东师范大学学报(教育科学版)》第 9 期。

李波, 2018:《教育补习真的有效吗?——基于学业表现和非认知发展的视角》,《北京社会科学》第 6 期。

李佳丽、胡咏梅, 2017:《谁从影子教育中获益?——兼论影子教育对教育结果均等化的影响》,《教育与经济》第 2 期。

李佳丽, 2018:《参加课外补习对西部农村学生的影响效应研究——基于面板数据的固定效应分析》,《基础教育》第 1 期。

庞晓鹏、严如贺、聂景春、罗斯高、张林秀、史耀疆、庞小冬, 2017:《农村小学生可以通过课外补习提高学习成绩吗》,《教育经济评论》第 2 期。

孙伦轩、唐晶晶, 2019:《课外补习的有效性——基于中国教育追踪调查的估计》,《北京大学教育评论》第 1 期。

魏易、薛海平, 2021:《校外培训机构治理是否有效?基于 2017—2019 年中国教育财政家庭调查数据的分析》,《教育科学研究》第 6 期。

薛海平、王东、巫锡炜, 2014:《课外补习对义务教育留守儿童学业成绩的影响研究》,《北京大学教育评论》第 3 期。

薛海平、赵阳, 2020:《高中生参加课外补习有助于考大学吗》,《华东师范大学学报(教育科学版)》第 5 期。

薛海平、李静, 2016:《家庭资本、影子教育与社会再生产》,《教育经济评论》第 4 期。

薛海平, 2015:《从学校教育到影子教育:教育竞争与社会再生产》,《北京大学教育评论》第 3 期。

薛海平, 2016:《课外补习、学习成绩与社会再生产》,《教育与经济》第 2 期。

杨莉, 2020:《班级同伴视角下课外补习效果的外溢性研究——基于 CEPS 数据的实证分析》,《教育经济评论》第 6 期。

张冰、程天君, 2019:《新中国成立以来学生“减负”历程的回顾与反思》,《教育科学》第 6 期。

Baker, D., M. Akiba, G. Tendre, A. Wiseman, J. Goldstein, 2001, “Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement”, *Educational Evaluation and Policy Analysis*, 23 (1): 1—17.

Bray, M. and P. Kwok, 2003, “Demand for Private Supplementary Tutoring: Conceptual Considerations, and Socio-Economic Patterns in Hong Kong”, *Economics of Education Review*, 22(6): 611—620.

Dang, H. A., 2007, “The Determinants and Impact of Private Tutoring Classes in

- Vietnam”, *Economics of Education Review*, 26(6): 683–698.
- Enrich, S., 2014, “Effects of Investments in Out-of-school Education in Germany and Japan”, *Contemporary Japan*, 26(1): 71–102.
- Heckman, J. J., H. Ichimura, P. E. Todd, 1997, “Matching as an Econometric Evaluation Estimator; Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies*, 64(4): 605–654.
- Kim, S., and J. H. Lee, 2010, “Private Tutoring and Demand for Education in South Korea”, *Economic Development and Cultural Change*, 58(2): 259–296.
- Stevenson, D. L., and D. P. Baker, 1992, “Shadow Education and Allocation in Formal Schooling; Transition to University in Japan”, *American Journal of Sociology*, 97(6): 1639–1657.
- Suryadarma, D., A. Suryahadi, S. Sumarto, and F. Rogers, 2006, “Improving Student Performance in Public Primary Schools in Developing Countries: Evidence from Indonesia”, *Education Economics*, 14(4): 401–429.
- Zhang, S., 2013, “Does Private Tutoring Improve Students’ National College Entrance Exam performance? A Case Study from Jinan, China”, *Economics of Education Review*, 32(1): 1–28.

The Influence of Private Tutoring on Academic Performance of junior high school students: Based on PSM-DID Method Estimation

GUAN Zhen

(Business School/ Capital Institute for Economics of Education, Beijing Normal University)

Abstract: Based on the China Education Panel Survey (CEPS) baseline and follow-up data, using the PSM-DID method, the impact of junior high school students’ participation in private tutoring on academic performance and the heterogeneity among different student groups were investigated. The main findings are as follows: (1) Parents’ educational level, family economic level and academic performance significantly increase the probability of students participating in math and English tutoring; but family economic level and academic performance have no significant impact on the probability of students participating in Chinese tutoring. (2) Chinese, mathematics and English tutoring can significantly improve students’ academic performance. English has the largest tuition effect, and the tutoring effect of Chinese and mathematics is similar. (3) For students with excellent academic performance, the effect of tutoring in each subject is greater. Therefore, tutoring in various disciplines will widen the gap in performance among students, which is not conducive to the balanced development of education.

Key words: extracurricular tutoring; academic performance; educational equity; difference-in-difference

(责任编辑: 刘泽云 责任校对: 刘泽云 胡咏梅)