

课外补习的有效性

——基于中国教育追踪调查的估计

孙伦轩 唐晶晶

(天津师范大学 教育科学学院, 天津 300387)

(陕西师范大学 教育实验经济研究所, 西安 710119)

摘要 课外补习能否有效提升学业成绩一直是研究和实践关注的焦点。本文基于中国教育追踪调查(CEPS)数据,使用双重差分倾向得分匹配法(PSM-DID)有效控制内生性问题,对课外补习的有效性进行了实证研究。研究发现:在总样本上,课外补习并不是提升学业成绩的“快车道”,甚至显著地降低了初中生的标准总成绩;与此同时,课外补习具有明显的“安慰剂效应”,显著地减少了初中生感到沮丧、悲伤、生活没有意思和不快乐等负面情绪的频率。进一步的分样本研究发现,“补习无效”主要存在于男性、农村和父母职业为非精英的弱势亚群体中,而“安慰剂效应”也常常在这些群体中同时出现。上述发现说明,对于优质教育机会不足的弱势群体学生来说,参与课外补习的根本动力可能并不在于学业成绩上的回报,更多的是一种心理和情感上的安慰。

关键词 课外补习;学业成绩;安慰剂效应;PSM-DID

中图分类号:G40-054 文献标识码:A 文章编号:1671-9468(2019)01-0123-19

DOI:10.12088/pku1671-9468.201901009

一、问题提出与文献回顾

教育是个体实现向上社会流动和防止向下社会流动的重要通道,因此父母们总是千方百计地给子女寻找“快车道”,课外补习就是其中之一。中国教育学会针对 2016 年国内课外补习的报告表明:有超过 1.37 亿青少年参与其中,并形成了超 8000 亿人民币的行业规模^[1]。小学到高中各教育阶段学生参加课外补习的比例分别达到 74%、66%和 53%^[2]。与学校教育推崇“公共理性”不同,课外补习市场奉行“经济理性”。在市场供需的基本逻辑当中,对补习机会的获

作者简介:孙伦轩,男,天津师范大学教育科学学院讲师,博士。

唐晶晶,女,陕西师范大学教育实验经济研究所硕士研究生。

得具有明显的阶层差异。有学者认为,课外补习已经成为学校教育之外另一种教育社会再生产的新机制,对教育公平和正常的社会流动形成了严重挑战^[3]。与此呼应,课外补习能否对学业成绩产生影响以及是否造成新的教育不平等,成为教育社会学关注的核心议题之一。比较一致的观点是,由于课外补习需要较高的经济资本,家庭社会经济地位较高的学生占有更多的补习机会。补习机会更多向发达地区、质量较高学校的学生聚集;在区域层面,参加课外补习的学生占比从高到低排序为:东部、中部、西部^[4]。基于九个经济转型期国家的研究也发现,城镇学生参与补习的比例比农村学生平均高 7.5%^[5]。对于农村学生来说,课外补习为弥补其与城市学生的竞争劣势带来希望,但大部分农村家庭都无法负担课外补习需要的额外费用^[6]。

值得注意的是,对于课外补习能否实质性提升学生的学业成绩,国内外研究都未能达成一致结论。例如,党海英(Hai-Anh Dang)以越南政府规定的课外补习费用为工具变量的研究发现,课外补习支出对越南小学生和初中生的成绩排名均有积极影响^[7];米斯克(Christoph Mischo)等人针对韩国的研究发现,课外补习能够提高韩国学生的数学成绩^[8]。然而,国外发现“补习无用”的研究也频繁出现。李德贤(Deockhyun Ryu)和康昌辉(Changhui Kang)发现,韩国初中生课外补习支出的增加对提高学生学习成绩并不明显,增加 10% 的支出只能提高 0.03 个标准差或 1.1% 的测试成绩^[9];苏亚达马拉(Gail Sunderman)等人针对印度尼西亚的研究发现,课外补习对小学四年级学生的成绩并无影响^[10];苏永勇(Soo-yong Byun)使用倾向值匹配法发现,韩国初中生私人家教、网络课程等课外补习法对学业成绩不发生作用(专业补习学校除外)^[11]。针对尼泊尔、埃及等国家的研究也有类似发现^{[12][13]}。中国的经验研究同样结论各异。一些研究认为,课外补习对学生成绩具有显著的促进作用。雷万鹏的研究发现,成绩越差的高中生参加课外补习的可能性更大且补习费用增长越快,这意味着课外补习市场弥补了正规学校教育的不足,对低成就学生具有公平效应^[14]。李佳丽和胡咏梅的研究证实了这一结论,即课外补习能够帮助学习困难的学生缩短与其他同学的课业差距^[15]。但是,也有研究发现相反的结果。张羽针对北京市某初中的研究发现,小学一年级参加语文或者数学课外补习,会对初中相应学科的初始成绩产生消极影响。小学高年级参与补习,能在一定程度上提高学生初中的初始成绩,却对初中三年的学业增长速度有消极影响^[16]。薛海平和丁小浩的研究表明,城市地区的私人课外补习扩大了教育不平等和社会分层^[17]。

无法准确识别课外补习对学业成就的因果效应是造成上述发现相互矛盾的重要原因之一。现有研究通常的做法是,比较参加补习和未补习学生在学业成就上的差异。众所周知,补习机会并不是随机分配的,受到个人特质以及家

庭背景等诸多因素的影响,也就是说参加补习(干预组)和未补习(控制组)的两组学生在不少特征变量上存在差异。这些差异也会影响学生的学业成就,从而产生内生性问题,这让研究者难以得到可靠的因果推断。通常情况下,断点回归(Regression Discontinuity,简称 RD)和工具变量(Instrumental Variable,简称 IV)是截面数据中解决上述内生性问题的可靠方法。但是,现实中很难找到可以对补习机会产生有效影响的外生断点,因此目前为止尚未见到断点回归在课外补习中的运用。同样,要想找到一个与内生变量强相关却与异质性残差无关的工具变量也非常有难度。例如张羽使用学生家庭与最近补习机构之间的距离以及学生同伴群体中参加补习的人数作为工具变量,但这两个变量仅能解决个体层面的内生性问题,却无法解决制度层面的内生性问题,因为他们皆为学生个体层面的变量^[18]。退而求其次,有研究者使用倾向值匹配法(P propensity Score Matching,简称 PSM)来控制样本选择性问题。然而,倾向值匹配虽然可以解决可观测的异质性问题,对不可观测的异质性却无能为力。此外,倾向值匹配也无法解决逆向因果的问题。已有研究表明,学生学业成绩的优劣很大程度上决定学生是否参加课外补习的概率^[19]。一般情况下,基于追踪数据的倍差法(Difference in Difference,简称 DID)是解决不可观测异质性和逆向因果的理想设计^[20]。它根据干预实施前和实施后的干预组和控制组指标值的差值,对两组差值的均值进行比较,实现了对两组的基线差异以及外部因素的控制,进而评估干预措施的净效应。但是,倍差法必须满足样本选择的随机性假设和共同趋势假设,加之追踪数据的获取非常难,因此在这一领域少有严格使用倍差法设计的研究。

现有研究除了在方法层面仍有改进空间之外,在事实逻辑的层面也需要进一步理顺。如果课外补习对学生的学业成绩没有影响甚至有显著的负向影响,那么家长和学生们为何会对课外补习趋之若鹜,进而形成如此巨大的补习市场?现有研究尚未给出一个合理且可信的回答。有学者认为,课外补习虽然对学生学业成绩的积极影响有限,但却对大部分学生的学业情绪起到“安慰剂”的作用。刘东芝的调查显示,尽管有 72.5% 的学生认为“以自己目前的学习情况,没有补习的必要”,但同时有 62.5% 的学生因为“大家都补,不补怕被别人超过”,30.1% 的学生因为“大家都补,不补怕跟不上”而选择继续补习。此外,学业导向的课外补习大多会提前学习到课堂的教学内容,这种准备状态让学生再次学习到相近知识时产生了熟悉感和胜任感,进而减少了学生的学业紧张和焦虑。大部分的学生对上补习班的感觉是“高兴”“快乐”和“学习后心情不错”,只有很少的学生认为上补习班后“不高兴”“很烦”和“心情很糟糕”^[21]。可见,课外补习的象征性意义及其准备性的补习安排确实可能导致学生对学业的情绪变化。在课外补习不能实质性提升学业成就的前提下,从课外补习中获得一种满

足感可能是形成大面积补习潮的根本动力之一。

综上,课外补习能否提升学生学业成绩?现有研究仍然结论不定、方法存疑或逻辑不顺。本文的创新之处在于:首先,借助于中国教育追踪调查(China Education Panel Survey,简称 CEPS)的两期追踪数据,运用倾向得分匹配法对可观测异质性予以处理,而针对不随时间变化的不可观测异质性残值则进一步使用倍差法,较为严格和稳健地估计了课外补习对学生成绩带来的净效应。其次,本文进一步检验了课外补习在初中生心理情绪层面带来的“安慰剂效应”,以期为学界理解大面积“补习潮”的成因提供新的视角和证据。本文发现:在总样本上,课外补习并不是提升学业成绩的“快车道”,甚至显著地降低了初中生的标准总成绩。与此同时,课外补习具有明显的“安慰剂效应”,显著地减少了初中生感到沮丧、悲伤、生活没有意思和不快乐等负性情绪的频率。进一步的分样本研究发现,“补习无效”主要存在于男性、农村和父母职业为非精英的弱势亚群体中,而“安慰剂效应”也常常在这些群体中同时出现。

二、数据来源与变量设计

(一) 数据来源

本文使用的数据来源于中国教育追踪调查(CEPS)数据。^① CEPS 由中国人民大学中国调查与数据中心设计并实施,是国内首个从初中阶段开始、针对青少年学生群体的大规模全国性追踪调查项目,因此本文的分析结果在全国具有可推广性。CEPS 使用多阶段概率与规模成比例(PPS)抽样,共有 112 所学校、438 个班级的约 2 万名学生参与了 2013—2014 年的基线调查。2014—2015 学年的追访对象为基线调查时初中一年级的全部 10279 名学生。在对重要数据变量缺失的样本进行剔除后,共有 6808 名学生进入有效样本。

(二) 变量设计

1. 自变量

为了尽可能广泛且稳健地考察课外补习带来的影响,本文分别将整体补习和单科补习作为核心自变量,包括参加过补习(是=1,否=0)、语文补习(是=1,否=0)、数学补习(是=1,否=0)、英语补习(是=1,否=0)。

2. 因变量

本文主要考察课外补习对学生学业成绩和心理情绪方面的影响。在学业

^① 该调查于 2013—2014 学年实施了基线调查,并于 2014—2015 学年进行了第二轮追访。调查详情(如抽样设计、调查手册、调查问卷、项目实施等)可参见 <http://ceps.ruc.edu.cn>。

成绩方面,CEPS收集了学生在2013年和2014年秋季学期数学、语文和英语期中考试成绩(学校提供),本文将其按照班级进行标准化:标准成绩=(原始成绩-班级最低成绩)/(班级最高成绩-班级最低成绩)。接下来,将语文、数学、英语的平均分处理为标准化总成绩:标准化总成绩=(标准化语文成绩+标准化数学成绩+标准化英语成绩)/3。考虑到各省市各学校的教材、考试难易程度及具体考核方法的差异可能导致标准化总成绩的有偏估计,故本文还考虑使用学生父母对其在班级成绩排名的评估作为学业成绩的另一个操作化指标,并将两个分析结果进行比较。CEPS调查询问学生父母:您估计孩子目前的成绩在班里位于哪一水平?回答选项依次为:1=不好,2=中下,3=中等,4=中上,5=很好;采用李克特五点计分法,可视为连续等距变量。最后,为了得到更为稳健的估计,本文还使用语文、数学和英语三门主课的自评学习难度作为因变量进行稳健性检验。CEPS调查对学生的数学、语文和英语学科的难度感知的三个题目是:你目前觉得数学/语文/英语课程学习起来吃力吗?回答选项均为:1=特别吃力,2=有点吃力,3=不是很吃力,4=一点也不吃力;本文均将其进行反向计分。在心理情绪方面,CEPS调查询问学生:在过去一周内,你是否有以下感觉?四个子问题分别是:沮丧、不快乐、生活没意思、悲伤;回答选项为1=从不,2=很少,3=有时,4=经常,5=总是;采用李克特五点计分法,可视为连续等距变量。

3. 协变量

协变量主要是样本学生不因是否补习而改变的个体、家庭及学校的特征变量,主要包括基线调查时学生的性别、是否寄宿、民族、户口类型、家庭经济层次、父母受教育年限、父母工作类型、是否公立学校、是否乡镇学校以及学校生均拨款层次。其中,性别(0=女,1=男)、是否寄宿(0=非寄宿生,1=寄宿生)、民族(0=汉族,1=少数民族)、户口类型(0=城镇户口,1=农村户口)、是否公立学校(0=否,1=是)、是否乡镇学校(0=否,1=是)为二分变量;家庭经济层次(0=困难、1=中等、2=富裕)为虚拟变量;生均拨款按照前25%、25%~50%、50%~75%、后25%分成四个档次,形成一个取值为1~4的连续变量;父母受教育程度仅使用父母双方受教育程度较高一方的信息,以受教育年限对其赋值^①;父母职业类型中,将管理类(国家机关、事业单位、企业的管理人员)和技术类(教师、工程师、医生、律师等)职业归为精英阶层,其他职业归为非精英阶层,并以父母双方得分较高一方的信息来表示父母职业类型^[22]。

^① 赋值原则:文盲=0年,小学毕业=6年,初中毕业=9年,中专/技校=11年,职业高中=11年,高中=12年,大学专科=15年,大学本科=16年,研究生及以上=19年。

表 1 变量的描述性统计

| | 变量 | 样本量 | 均值 | 标准差 | 最小值 | 最大值 | |
|--------|---------|--------|---------|--------|--------|-----|---|
| 自变量 | 整体补习 | 6,808 | 0.4852 | 0.4998 | 0 | 1 | |
| | 数学补习 | 6,808 | 0.2337 | 0.4232 | 0 | 1 | |
| | 语文补习 | 6,808 | 0.1222 | 0.3276 | 0 | 1 | |
| | 英语补习 | 6,808 | 0.2609 | 0.4391 | 0 | 1 | |
| 因变量 | 基期 | 父母评估排名 | 6,808 | 3.1661 | 1.0270 | 1 | 5 |
| | | 标准化总成绩 | 6,808 | 0.6705 | 0.2060 | 0 | 1 |
| | | 数学成绩 | 6,808 | 0.6596 | 0.2576 | 0 | 1 |
| | | 语文成绩 | 6,808 | 0.6538 | 0.2238 | 0 | 1 |
| | | 英语成绩 | 6,808 | 0.6980 | 0.2550 | 0 | 1 |
| | | 数学难度 | 6,808 | 2.4634 | 0.9063 | 1 | 4 |
| | | 语文难度 | 6,808 | 2.7288 | 0.7877 | 1 | 4 |
| | | 英语难度 | 6,808 | 2.6062 | 0.9592 | 1 | 4 |
| | | 沮丧 | 6,808 | 2.1576 | 0.9533 | 1 | 5 |
| | | 不快乐 | 6,808 | 2.2081 | 1.0056 | 1 | 5 |
| | | 生活没有意思 | 6,808 | 1.6561 | 1.0134 | 1 | 5 |
| | 悲伤 | 6,808 | 1.9802 | 0.9976 | 1 | 5 | |
| | 评估 | 父母评估排名 | 6,808 | 3.1535 | 1.0470 | 1 | 5 |
| | | 标准化总成绩 | 6,808 | 0.6410 | 0.2229 | 0 | 1 |
| | | 数学成绩 | 6,808 | 0.6416 | 0.2776 | 0 | 1 |
| | | 语文成绩 | 6,808 | 0.6651 | 0.2248 | 0 | 1 |
| | | 英语成绩 | 6,808 | 0.6162 | 0.2799 | 0 | 1 |
| | | 数学难度 | 6,808 | 2.5198 | 0.8714 | 1 | 4 |
| | | 语文难度 | 6,808 | 2.8638 | 0.7504 | 1 | 4 |
| | | 英语难度 | 6,808 | 2.3995 | 0.9831 | 1 | 4 |
| | | 沮丧 | 6,808 | 2.3001 | 1.0477 | 1 | 5 |
| | | 不快乐 | 6,808 | 2.2828 | 1.0577 | 1 | 5 |
| 生活没有意思 | | 6,808 | 1.8988 | 1.0934 | 1 | 5 | |
| 悲伤 | 6,808 | 2.0958 | 1.0527 | 1 | 5 | | |
| 协变量 | 男生 | 6,808 | 0.5116 | 0.4999 | 0 | 1 | |
| | 寄宿生 | 6,808 | 0.2991 | 0.4579 | 0 | 1 | |
| | 少数民族 | 6,808 | 0.0736 | 0.2611 | 0 | 1 | |
| | 农村户口 | 6,808 | 0.5073 | 0.5000 | 0 | 1 | |
| | 家庭经济层次 | 6,808 | 0.8575 | 0.4869 | 0 | 2 | |
| | 父母受教育年限 | 6,808 | 10.7854 | 3.1431 | 0 | 19 | |
| | 父母的职业类型 | 6,808 | 0.2522 | 0.4343 | 0 | 1 | |
| | 公立学校 | 6,808 | 0.9385 | 0.2403 | 0 | 1 | |
| | 乡镇学校 | 6,808 | 0.3327 | 0.4712 | 0 | 1 | |
| | 生均拨款层次 | 6,808 | 2.0091 | 1.2457 | 1 | 4 | |

三、估计方法与分析步骤

(一) 双重差分法

双重差分法在当前教育评估领域运用广泛。该方法的基本程序为:将样本分为干预组和控制组,以控制组的结果作为干预组的反事实结果,并对两个结果的差异进行计算,进而得到因果效应。与其他方法相比,其最大优势是可以利用追踪数据来控制不可观测变量的影响,尤其是可以控制不随时间改变和随时间同步变化的因素的影响^[23]。本文的干预组为参加课外补习的学生,控制组为未参加课外补习的学生,将控制组的学业成绩及心理情绪作为干预组学生的学业成绩及心理情绪的反事实结果,然后计算干预组学生学业成绩及心理情绪的事实结果与反事实结果之差,进而得到参加课外补习对学生学业成绩及心理情绪的影响效应,即平均处理效应(Average Treatment Effect on the Treated, 简称 ATT)。计算公式如下:

$$ATT_{DID} = E(Y_{t_1}^T - Y_{t_0}^T | D = 1) - E(Y_{t_1}^C - Y_{t_0}^C | D = 0) \quad (1)$$

其中, E 为期望算子; D 为参加课外补习的虚拟变量,干预组和控制组分别赋值为 1 和 0; T 、 C 分别表示干预组和控制组; t_0 、 t_1 分别为基线即学生七年级的时点和追踪时即学生八年级的时点; Y_{t_0} 、 Y_{t_1} 分别表示学生七年级时和八年级时的结果变量(即学业成绩和心理情绪)。式(1)右边两项分别为干预组和控制组在两次调查时自身学业成绩及心理情绪的差分,目的是消除两组样本自身的变化趋势,两组样本自身差分的差分就是参加课外补习对学业成绩及心理情绪的影响效应。

(二) 倾向得分匹配法

双重差分法必须满足严格的前提条件:即样本选择的随机性假设和共同趋势假设。但参加课外补习与未参加课外补习的学生本身就在个人、家庭以及学校等方面存在差别,并非随机分配。因此,简单地把未参加课外补习学生的学业成绩及心理情绪作为参加课外补习学生学业成绩及心理情绪的反事实结果会出现样本选择偏差问题,需要在做双重差分之前选取与参加课外补习学生特征相似的样本作为控制组。

为了克服选择性偏差等内生性问题对因果推断造成的影响,本文在评估参加课外补习对初中生学业成绩及心理情绪的影响前使用倾向值匹配法将干预组和控制组进行匹配。倾向值匹配法是近些年来在控制选择性偏差中较为有效的一种因果推论方法,其基本思想是:依据反事实框架(Counterfactual Framework),设置一个反事实组(即控制组),使其与干预组的可观测协变量分

布尽量趋同^[24]。倾向值是指某个单位在给定可观测协变量的条件下落入干预组的概率,研究者根据估计的倾向值将干预组成员与控制组类似成员匹配以获得“平衡”的、类似于随机实验的设计。匹配后的样本满足条件独立分布假设,通过比较最终匹配的两组在因变量上的平均差异,得到“平均干预效应”(Average Treatment Effect,简称 ATE)。假定有因变量 Y 、干预变量 $D(D=1$ 为干预组 $T, D=0$ 为控制组 $C)$ 和一组可观测的协变量 X 。

倾向值匹配分析的第一步是估计倾向值:

$$P(x) = \Pr[D = 1 | X] \quad (2)$$

在满足无混淆和存在共同支持假定时,基于倾向值的匹配可以实现与基于协变量 X 的匹配一样好的效果。第二步是基于估计的倾向值将干预组和控制组匹配起来,常见的匹配方法包括邻近匹配、半径匹配和核匹配。第三步是利用匹配样本估计因果效应。干预组的平均干预效应(Average Treatment Effect on the Treated, ATT)可定义为:

$$\begin{aligned} \text{ATT}_{\text{PSM}} &= E\{Y_i(1) - Y_i(0) | D = 1\} = E\{Y_i(1) - Y_i(0) | D = 1, p_i(x)\} \\ &= E\{Y_i(1) | D = 1, p_i(x)\} - E\{Y_i(0) | D = 1, p_i(x)\} \end{aligned} \quad (3)$$

(三) 双重差分倾向值匹配法

尽管倾向值匹配法可以克服可观测变量的样本选择偏差问题,但是它只能基于可观测到的变量对两组样本进行匹配。换句话说,该方法假定初中生是否参加课外补习完全取决于可观测变量,不可观测变量则不会对他们的选择产生任何影响。因此,该方法不能纠正不可观测变量对初中生是否参加课外补习的影响,简单地利用倾向值匹配法估计出的平均处理效应仍然是有偏的。双重差分法可以克服不可观测变量的影响,正好可以弥补倾向值匹配法在这一方面的不足。为此,赫克曼(J. J. Heckman)等提出了双重差分倾向值匹配法(PSM-DID),可以充分利用双重差分法和倾向值匹配法的各自优点,同时克服不可观测变量和可观测变量对样本选择的影响^{[25][26]}。

因此,本文基于 CEPS 两期的数据,使用 PSM-DID 方法同时控制不可观测变量和可观测变量对初中生是否参加课外补习决策的影响,有效地解决样本选择偏差问题,进而得到参加课外补习对初中生学业成绩及心理情绪的平均干预效应(ATT):

$$\text{ATT}_{\text{PSM-DID}} = E(Y_{t_1}^T - Y_{t_0}^T | X_{t_0}, D = 1) - E(Y_{t_1}^C - Y_{t_0}^C | X_{t_0}, D = 0) \quad (4)$$

其中, X_{t_0} 为协变量,表示两组样本七年级时可能影响他们是否参加课外补习的一组特征变量,这些协变量不仅影响他们是否参加课外补习,还可能影响他们的学业成绩和心理情绪。其他变量与上文相同,式(4)为估计参加课外补习对初中生学业成绩及心理情绪影响的基准模型。

四、实证分析

(一) 倾向值得分匹配

本文首先以初中生是否参加补习(包括整体课外补习和各单科补习)为因变量建立倾向值预测模型,得到学生参加补习的 logit 模型回归结果,如表 2 所示。匹配前四个 logit 模型的回归结果均拟合较好。^① 四个模型均通过了 ROC 检验,表示模型选择的协变量对学生是否参加课外补习有比较强的解释力,由此就可以得到初中生参加课外补习的预测概率,并据此对干预组和控制组进行核匹配。

(二) 样本匹配质量检验

1. 平衡性检验

PSM 模型消除样本选择性偏差的效果取决于匹配过程中是否满足条件独立性假设及共同支撑域两个前提假设。如果匹配未能满足前提假设,则表示协变量或匹配方法的选择不恰当,PMS 估计结果是存在偏差的。本文分别以参加课外补习、参加数学补习、参加语文补习、参加英语补习为自变量进行匹配,并采用 T 检验对匹配前后各变量是否存在显著的组间差异进行了检验,结果如表 2 所示。具体来看,匹配后干预组和控制组协变量的标准化偏差削减后最大值为 8.5%,符合匹配后两组协变量的标准化偏差不超过 10%的标准。与此同时, T 检验的结果都不拒绝干预组与控制组无系统差异的原假设,这意味着两组学生的个体差异得到了有效控制,各协变量的均衡效果较好。此外,四组样本匹配后的倾向值模型的 R^2 均大幅度减小,分别为 0.001、0.000、0.002、0.001^②。由解释变量联合显著性检验较大的 p 值(分别为 0.877、1.000、0.949、0.987)可知,匹配后的协变量和倾向得分在干预组和控制组之间的分布是一致的,满足整体平衡性条件。

① 因变量为是否参加任何补习时,LRchi2=1153.55, $p=0.0000$,Pseudo $R^2=0.1223$;因变量为是否参加数学补习时,LRchi2=495.14, $p=0.0000$,Pseudo $R^2=0.0669$;因变量为参加语文补习时,LRchi2=180.83, $p=0.0000$,Pseudo $R^2=0.0358$;因变量为参加英语补习时,LRchi2=812.58, $p=0.0000$,Pseudo $R^2=0.1040$ 。

② 模型 R^2 值越小,表示协变量的联合显著性越差,匹配后的干预组和控制组之间的差异就越小,匹配质量就越高。

表 2 匹配前后协变量均值偏差及组间差异检验

| 协变量 | 是否匹配 | 补习 | | 数学补习 | | 语文补习 | | 英语补习 | |
|--------------|------|---------|-----------|--------|-----------|--------|-----------|--------|-----------|
| | | 偏差% | $p > t $ | 偏差% | $p > t $ | 偏差% | $p > t $ | 偏差% | $p > t $ |
| 男性 | 匹配前 | 9.6 | 0.000 | -6.1 | 0.034 | 3.4 | 0.361 | -4.1 | 0.142 |
| | 匹配后 | -0.4 | 0.870 | -0.5 | 0.895 | 1.4 | 0.782 | 0 | 1.000 |
| 寄宿生 | 匹配前 | 46.2 | 0.000 | -40.8 | 0.000 | -15 | 0.000 | -48.9 | 0.000 |
| | 匹配后 | 0.2 | 0.946 | -1.9 | 0.546 | -3.1 | 0.513 | -2.8 | 0.323 |
| 少数民族 | 匹配前 | 5.4 | 0.026 | -7.8 | 0.008 | -7.3 | 0.061 | -5 | 0.078 |
| | 匹配后 | 2.7 | 0.275 | 0.2 | 0.960 | -0.9 | 0.845 | -0.4 | 0.890 |
| 农村户口 | 匹配前 | 55.3 | 0.000 | -42.7 | 0.000 | -28.9 | 0.000 | -59 | 0.000 |
| | 匹配后 | -0.9 | 0.704 | -1.7 | 0.617 | -5.7 | 0.240 | -0.5 | 0.872 |
| 家庭经济层次 | 匹配前 | -37.7 | 0.000 | 28.7 | 0.000 | 26.6 | 0.000 | 34.4 | 0.000 |
| | 匹配后 | -2 | 0.449 | 1.3 | 0.704 | 5.4 | 0.256 | 1.8 | 0.570 |
| 父母受教育年限 | 匹配前 | -65.8 | 0.000 | 46.5 | 0.000 | 41.7 | 0.000 | 63 | 0.000 |
| | 匹配后 | -0.5 | 0.822 | 2.7 | 0.471 | 8.5 | 0.096 | 0.6 | 0.869 |
| 父母的职业类型 | 匹配前 | -52.4 | 0.000 | 34.5 | 0.000 | 33.1 | 0.000 | 47 | 0.000 |
| | 匹配后 | -1 | 0.612 | 2.4 | 0.526 | 6.5 | 0.210 | -1.4 | 0.706 |
| 公立学校 | 匹配前 | -8.5 | 0.000 | 4.5 | 0.124 | 2.4 | 0.517 | 15.4 | 0.000 |
| | 匹配后 | 3.2 | 0.236 | 0.3 | 0.941 | 0.8 | 0.871 | 0.9 | 0.753 |
| 乡镇学校 | 匹配前 | 56.2 | 0.000 | -43.9 | 0.000 | -29.8 | 0.000 | -54.3 | 0.000 |
| | 匹配后 | 1.7 | 0.524 | -1.5 | 0.635 | -5.1 | 0.271 | -2.5 | 0.385 |
| 生均拨款层次 | 匹配前 | -35.3 | 0.000 | 34.5 | 0.000 | 18.8 | 0.000 | 27.8 | 0.000 |
| | 匹配后 | -2.4 | 0.283 | 1.8 | 0.636 | 4 | 0.431 | -1.8 | 0.614 |
| Pseudo R^2 | 匹配前 | 0.122 | | 0.067 | | 0.036 | | 0.104 | |
| | 匹配后 | 0.001 | | 0.000 | | 0.002 | | 0.001 | |
| LR χ^2 | 匹配前 | 1153.55 | | 495.14 | | 180.83 | | 812.58 | |
| | 匹配后 | 5.21 | | 0.94 | | 3.96 | | 2.72 | |
| $P > \chi^2$ | 匹配前 | 0.000 | | 0.000 | | 0.000 | | 0.000 | |
| | 匹配后 | 0.877 | | 1.000 | | 0.949 | | 0.987 | |

2. 共同支撑检验

在估计平均处理效应之前,还需进行共同支撑检验,即检验倾向值在干预组和控制组中是否有足够大的重合区域即共同支撑区域,以确保倾向值匹配的有效性^[27]。四次匹配的倾向值的共同支撑分布如图 1 所示:落在共同支撑域外的样本量很小,倾向值在干预组和控制组中具有足够大的共同支撑域,可确保倾向值匹配的有效性和平均干预效应的准确性。

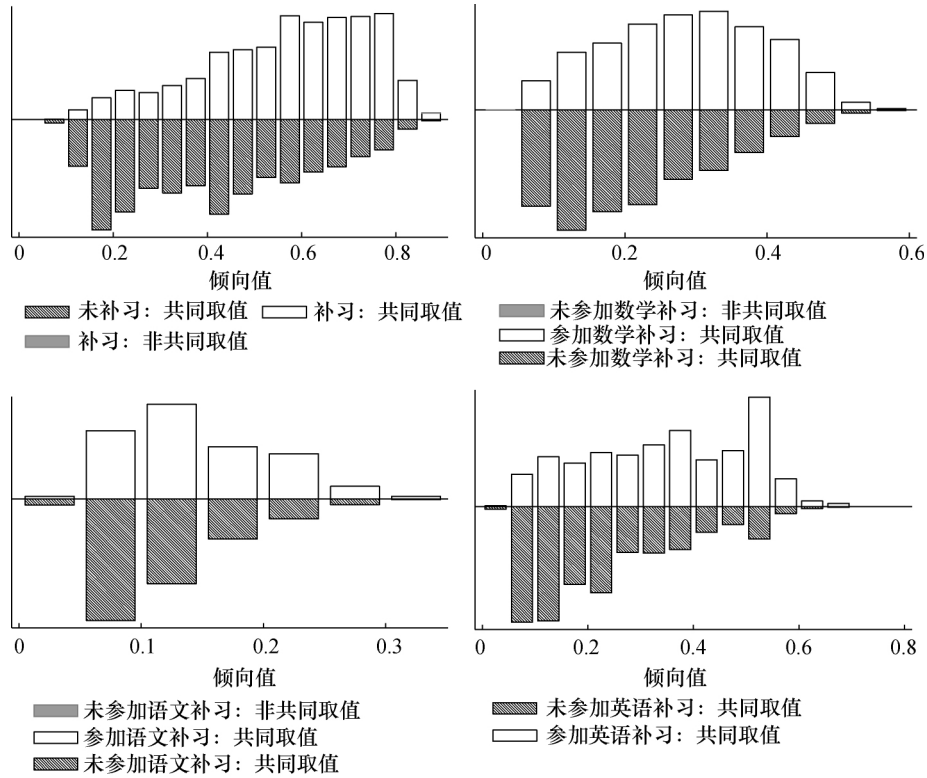


图 1 倾向指数拟合值的分布

(三) 课外补习对学业成绩的影响

由以上平衡性检验和共同支撑检验可知,匹配后的样本满足 PSM-DID 方法的前提假设条件(包括条件独立分布假设和共同支撑假设)。因此,本文根据基准模型式(4)分别以父母评估成绩、标准化总成绩、学科标准化成绩作为结果变量进行回归,得到课外补习对初中生学业成绩的平均干预效应。结果如表 5 所示,参加课外补习对初中生学业成绩几乎无显著影响。具体来看,参加课外补习之前,干预组学生的父母对其学业成绩评估的排名比控制组高 0.033;参加补习之后,干预组比控制组高 0.028,课外补习降低了干预组学生父母对其学业成绩评价的排名,但未能通过显著性检验。参加课外补习之前干预组的标准化总成绩比控制组高 0.014;参加课外补习之后,干预组的标准化总成绩比控制组高 0.004,参加课外补习对学生学业成绩总分的净效应约为 -0.010;在 0.1 的水平上通过了显著性检验。同理,数学补习对数学成绩的净效应为 0.010;语文补习对语文成绩的净效应为 0.008;英语补习对英语成绩的净效应为 0.001,虽都略有增长,但均未通过显著性检验。综上,无论是标准化成绩和父母评估排名,课外补习均未产生显著的提升作用,反而显著降低了学生的标准化总成绩。

表 3 课外补习对初中生学业成绩的平均干预效应

| | 七年级 | | | 八年级 | | | 平均处 理效应 |
|------------|-------|-------|----------|-------|-------|----------|------------|
| | 控制组 | 干预组 | 差分 | 控制组 | 干预组 | 差分 | |
| 父母评估成绩 | 3.110 | 3.143 | 0.033 | 3.110 | 3.138 | 0.028 | -0.005 |
| 标准误 | | | 0.030 | | | 0.029 | 0.023 |
| <i>t</i> 值 | | | 1.08 | | | 0.97 | 0.22 |
| $p > t $ | | | 0.283 | | | 0.22 | 0.828 |
| R^2 | | | | | | | 0.00 |
| 标准化总成绩 | 0.652 | 0.666 | 0.014 | 0.624 | 0.628 | 0.004 | -0.009 |
| 标准误 | | | 0.006 | | | 0.007 | 0.005 |
| <i>t</i> 值 | | | 2.13 | | | 0.66 | 1.87 |
| $p > t $ | | | 0.034** | | | 0.508 | 0.062* |
| R^2 | | | | | | | 0.01 |
| 数学成绩 | 0.652 | 0.648 | -0.004 | 0.641 | 0.647 | 0.006 | 0.010 |
| 标准误 | | | 0.011 | | | 0.011 | 0.011 |
| <i>t</i> 值 | | | -0.40 | | | 0.52 | 0.96 |
| $p > t $ | | | 0.687 | | | 0.600 | 0.340 |
| R^2 | | | | | | | 0.00 |
| 语文成绩 | 0.653 | 0.647 | -0.006 | 0.665 | 0.667 | 0.002 | 0.008 |
| 标准误 | | | 0.011 | | | 0.012 | 0.010 |
| <i>t</i> 值 | | | -0.55 | | | 0.14 | 0.76 |
| $p > t $ | | | 0.585 | | | 0.888 | 0.446 |
| R^2 | | | | | | | 0.00 |
| 英语成绩 | 0.710 | 0.746 | 0.037 | 0.640 | 0.676 | 0.035 | -0.001 |
| 标准误 | | | 0.011 | | | 0.012 | 0.009 |
| <i>t</i> 值 | | | 3.25 | | | 3.00 | 0.14 |
| $p > t $ | | | 0.001*** | | | 0.003*** | 0.888 |
| R^2 | | | | | | | 0.02 |

注: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

(四) 稳健性检验

为了进一步证实以上结论,本文用学生对各科课程的难度感知作为结果变量来对上述结果发现进行稳健性检验。理论上,学生对学科难度的自我评估一定程度上体现出其在本科目的学业表现,因此可以作为学业成绩的代理变量进行稳健性检验。结果如表 4 所示,参加数学和英语的课外补习对学生评价这两门学科的学习难度并无显著影响,而参加语文补习却使初中生认为语文学习

更难。具体来看,参加数学课外补习之前,干预组学生认为的数学难度比控制组高 0.032;参加数学课外补习之后,干预组比控制组高 0.050;课外补习增加了干预组学生心中数学学科的难度,但未能通过显著性检验。参加语文课外补习之前,干预组学生心中语文学科的难度比控制组低 0.017;参加补习之后,干预组比控制组高 0.055;语文补习对语文学科自评难度的净效应为 0.072,且在 0.05 的水平上通过显著性检验。同理,英语补习对英语学科自评难度的净效应为 -0.017,但并不显著。稳健性检验的结果继续支持了上述“补习无效论”。也就是说,初中生参加课外补习对其学业成绩没有正向影响,也不会降低学生对该课程难度的评价,甚至增加了对语文学科学习难度的认知。

表 4 课外补习对自评学业难度的平均干预效应

| | 七年级 | | | 八年级 | | | 平均干预效应 |
|------------|-------|-------|----------|-------|-------|----------|---------|
| | 控制组 | 干预组 | 差分 | 控制组 | 干预组 | 差分 | |
| 数学难度 | 2.509 | 2.541 | 0.032 | 2.582 | 2.632 | 0.050 | 0.017 |
| 标准误 | | | 0.031 | | | 0.032 | 0.031 |
| <i>t</i> 值 | | | 1.01 | | | 1.53 | 0.55 |
| $p > t $ | | | 0.314 | | | 0.128 | 0.585 |
| R^2 | | | | | | | 0.000 |
| 语文难度 | 2.766 | 2.749 | -0.017 | 2.897 | 2.952 | 0.055 | 0.072 |
| 标准误 | | | 0.029 | | | 0.038 | 0.034 |
| <i>t</i> 值 | | | -0.58 | | | 1.45 | 2.11 |
| $p > t $ | | | 0.564 | | | 0.150 | 0.036** |
| R^2 | | | | | | | 0.01 |
| 英语难度 | 2.649 | 2.869 | 0.220 | 2.519 | 2.722 | 0.203 | -0.017 |
| 标准误 | | | 0.037 | | | 0.035 | 0.033 |
| <i>t</i> 值 | | | 5.90 | | | 5.74 | 0.50 |
| $p > t $ | | | 0.000*** | | | 0.000*** | 0.616 |
| R^2 | | | | | | | 0.02 |

注: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$ 。

(五) “安慰剂效应”

本文用初中生心理情绪的四个子指标来检测课外补习可能带来的“安慰剂效应”,结果如表 5 所示。课外补习显著地提升了初中生的心理情绪水平,减少了学生感到沮丧、不快乐、生活没意思和悲伤等负性情绪的频率。具体来看,参加课外补习之前,干预组学生感到沮丧的频率比控制组高 0.027;参加课外补习之后,干预组比控制组低 0.034;课外补习的净效应为 -0.061,在 0.05 的水平

显著。参加课外补习之前,干预组学生感到不快乐的频率比控制组高 0.038;参加课外补习之后,干预组比控制组低 0.019;课外补习的净效应为-0.057,在 0.1 的水平上显著。参加课外补习之前,干预组学生感到生活没有意思的频率比控制组高 0.094;参加课外补习之后,干预组比控制组高 0.029;课外补习的净效应为-0.065,在 0.05 的水平上显著。参加课外补习之前,干预组感到悲伤的频率比控制组高 0.018;参加课外补习之后,干预组比控制组低 0.067;课外补习的净效应为-0.085,在 0.01 的水平上显著。综上,课外补习对学生的心理状态产生了显著的积极影响,具有明显的“安慰剂效应”。

表 5 课外补习的“安慰剂效应”

| | 七年级 | | | 八年级 | | | 平均干预效应 |
|------------|-------|-------|----------|-------|-------|----------|----------|
| | 控制组 | 干预组 | 差分 | 控制组 | 干预组 | 差分 | |
| 沮丧 | 2.170 | 2.197 | 0.027 | 2.322 | 2.288 | -0.034 | -0.061 |
| 标准误 | | | 0.024 | | | 0.025 | 0.031 |
| <i>t</i> 值 | | | 1.14 | | | 1.34 | 1.97 |
| $p > t $ | | | 0.256 | | | 0.183 | 0.050** |
| R^2 | | | | | | | 0.00 |
| 不快乐 | 2.254 | 2.293 | 0.038 | 2.331 | 2.312 | -0.019 | -0.057 |
| 标准误 | | | 0.023 | | | 0.027 | 0.031 |
| <i>t</i> 值 | | | 1.68 | | | 0.69 | 1.81 |
| $p > t $ | | | 0.095* | | | 0.488 | 0.072* |
| R^2 | | | | | | | 0.00 |
| 生活没有意思 | 1.656 | 1.750 | 0.094 | 1.925 | 1.954 | 0.029 | -0.065 |
| 标准误 | | | 0.025 | | | 0.025 | 0.029 |
| <i>t</i> 值 | | | 3.83 | | | 1.15 | 2.29 |
| $p > t $ | | | 0.000*** | | | 0.251 | 0.023** |
| R^2 | | | | | | | 0.01 |
| 悲伤 | 2.033 | 2.051 | 0.018 | 2.175 | 2.108 | -0.067 | -0.085 |
| 标准误 | | | 0.025 | | | 0.025 | 0.029 |
| <i>t</i> 值 | | | 0.74 | | | 2.69 | 2.98 |
| $p > t $ | | | 0.462 | | | 0.008*** | 0.003*** |
| R^2 | | | | | | | 0.00 |

注: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

(六) 不同亚群体课外补习效果的异质性

上述关于“补习无效性”和“安慰剂效应”的发现是否存在于所有亚群体？本文需要进一步考虑不同亚群体补习效果的异质性。因为并非所有的学生补习效果都一样，而这有可能造成整体样本的估计结果不显著。因此，本文根据性别、城乡、生均经费和父母职业将总样本分为八个子样本，分析结果如表 6 所示。整体来看，接受课外补习的干预组在绝大部分成绩指标上未有显著增长，除了城市学生和生均经费前 50% 的学生的数学成绩以及农村学生的英语成绩。与此同时，课外补习确实显著减少了绝大部分学生感受到负性情绪的频率。但必须注意的是，课外补习对学生学业成绩和心理情绪的影响在不同亚群体之间确实存在异质性。

将总样本和分样本的分析结果比较来看，在总成绩方面，课外补习在总样本中的影响是显著为负的，但分样本的结果显示这种负向影响主要存在于男性、农村地区和父母职业为非精英阶层的弱势群体中，对另外一些亚群体的影响则不显著。在数学成绩方面，课外补习在总样本中的影响不显著，在分样本中则对城市和生均经费前 50% 的学生产生了显著的积极影响，而对农村地区和生均经费后 50% 的学生产生不显著的负向影响，这可能是导致对总样本影响不显著的原因。在父母评估成绩、语文成绩和英语成绩方面，课外补习的影响在绝大多数样本中未通过显著性检验，除了农村样本的英语成绩。课外补习对农村学生的英语成绩产生了显著的积极影响，这可能是由于农村地区英语教育质量较低导致的。在心理情绪方面，课外补习显著减少了总样本中各种负性情绪的频率，但分样本的结果显示，这种积极的影响主要存在于男性、农村地区、生均经费后 50% 和父母职业为非精英阶层的弱势学生群体中。虽然课外补习对其他优势学生群体的心理情绪有积极影响，但绝大多数都未通过显著性检验。

如果将学业成绩和心理情绪的结果联系起来看，可以发现“补习无效”现象与“安慰剂效应”常常在同样的子样本中出现。例如，课外补习显著地减少了男生的标准化总成绩，但同时显著地减少了他们感受到悲伤和生活没有意思的频率，而这种一致性并未在女性群体中发现。类似的，课外补习减少了父母职业为非精英阶层学生样本的标准化总成绩，但同时也减少了他们感受到各种负性情绪的频率。因此，本文推断，“补习无效论”主要存在于男性、农村地区和父母职业为非精英的弱势亚群体中，当这些亚群体无法通过课外补习实质性提升学业成绩的时候，课外补习在心理情绪层面带来的“安慰剂效应”则成为他们参与补习的主要动力。

表6 不同亚群体课外补习效果的异质性

| 结果变量 | 性别 | | 城乡 | | | 生均经费 | | | 父母职业 | |
|--------|----------------------|-------------------|---------------------|----------------------|----------------------|---------------------|-------------------|----------------------|------|--|
| | 男性 | 女性 | 城市 | 农村 | 后50% | 前50% | 精英 | 非精英 | | |
| 父母评估成绩 | 0.007 (0.034) | -0.018 (0.032) | 0.034 (0.034) | -0.020 (0.033) | -0.006 (0.034) | 0.004 (0.031) | -0.018 (0.054) | -0.001 (0.026) | | |
| 标准总成绩 | -0.013** (0.006) | -0.004 (0.006) | -0.004 (0.007) | -0.011* (0.006) | -0.006 (0.008) | -0.013** (0.005) | -0.002 (0.009) | -0.010** (0.005) | | |
| 数学成绩 | 0.011 (0.011) | 0.010 (0.013) | 0.027* (0.014) | -0.010 (0.010) | -0.012 (0.011) | 0.029* (0.015) | 0.020 (0.015) | 0.009 (0.010) | | |
| 语文成绩 | 0.011 (0.014) | 0.001 (0.011) | 0.009 (0.013) | 0.005 (0.015) | 0.008 (0.015) | 0.007 (0.014) | 0.004 (0.013) | 0.010 (0.012) | | |
| 英语成绩 | -0.005 (0.012) | 0.000 (0.009) | -0.017 (0.011) | 0.036*** (0.010) | 0.001 (0.014) | -0.002 (0.010) | -0.018 (0.011) | 0.011 (0.010) | | |
| 沮丧 | -0.075* (0.041) | -0.015 (0.042) | -0.051 (0.040) | -0.070* (0.039) | -0.093** (0.044) | -0.033 (0.043) | 0.010 (0.057) | -0.075** (0.033) | | |
| 不开心 | -0.075 (0.046) | -0.041 (0.039) | -0.065 (0.043) | -0.057 (0.041) | -0.072 (0.045) | -0.053 (0.043) | -0.065 (0.061) | -0.071** (0.035) | | |
| 生活没有意思 | -0.083** (0.041) | -0.048 (0.039) | -0.088** (0.044) | -0.065* (0.037) | -0.084* (0.043) | -0.070* (0.039) | -0.076 (0.062) | -0.074** (0.032) | | |
| 悲伤 | -0.116*** (0.042) | -0.045 (0.040) | -0.018 (0.041) | -0.118*** (0.038) | -0.138*** (0.041) | -0.022 (0.041) | -0.065 (0.056) | -0.086*** (0.032) | | |

注:括号内为标准误;* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

五、结论与讨论

课外补习是否导致了新的教育不平等?这一议题引起了社会和学界的广泛关注,具有十分重要的理论与实践意义。准确识别课外补习对学生学业成就的因果效应至关重要。本文基于中国教育追踪调查(CEPS)两期的数据,使用双重差分倾向值匹配法(PSM-DID)实证探索了课外补习对初中生学业成绩的影响。研究发现:在总样本上,无论是标准化总成绩,还是语文、数学、外语三门单科标准化成绩,乃至父母对孩子在校排名上的评估成绩,课外补习都未能带来的显著的正向影响,甚至显著地减少了标准化总成绩。这与张羽和李佳丽等人的发现是一致的^{[28][29]}。那么,是什么驱使学生在“补习无效”的情况下仍然对课外补习趋之若鹜呢?研究进一步发现,课外补习在总样本上给学生心理层面带来了较为稳健的“安慰剂效应”,即课外补习显著地减少了学生感到沮丧、不快乐、生活没有意思以及悲伤等负面情绪的频率,这可能是驱使大量学生参与课外补习的重要动力之一。

然而,上述发现并不意味着课外补习没有造成新的教育不平等。本文进一步将总样本拆分为八个子样本的分析发现,课外补习确实减少了某些学生的标准总成绩,但这些学生主要是男性、农村和父母职业为非精英阶层的群体,而对城市和生均经费前50%的学生影响不大。这说明课外补习可能让弱势群体受损,从而扩大现有的教育不平等。进一步探索不同子样本中课外补习的“安慰剂效应”可以发现,它主要存在于男生、农村地区、生均经费后50%和父母职业为非精英阶层的弱势群体中。综合来看,课外补习对弱势群体学生来说更多发挥的是一种情感价值,而非基于提升成绩的工具价值。这可能有两个原因:其一,弱势群体的课外补习效能更差。尤其是农村和生均经费较低地区的课外补习,质量堪忧。男生虽然在教育资源的占有上不是弱势群体,但男生在中国传统课堂中的表现一直处于弱势地位,近些年来广泛探讨的“男孩危机”就印证了这一点^{[30][31]}。其二,在农村和经济欠发达地区,许多父母的受教育程度较低,因此将教育孩子的希望寄托于补习机构。然而,他们缺乏辨别课外补习是否有效的能力,并且更有可能因常年外出务工而无法陪伴孩子。因此,这些弱势群体学生有可能接受低效或无效的课外补习,而从参加补习的象征性意义中获得优越感恰恰填补了父母陪伴较少带来的心理缺失。

上述发现具有较强的社会意义和政策意义。对于那些在各种生活机遇上本就处于劣势的弱势群体学生来说,真正驱使他们参与课外补习的动力,可能并不源于学业成绩上的回报,更多的是一种从众心理带来的担忧与恐慌。课外补习能够有效地缓解这种压力,是因为参加补习的学生可能会因为课外补习的象征性意义而感到一种相对心理优势。但也需要注意到,这种“安慰剂效应”可能会产生一种麻痹作用,让这些学生误认为自己已经掌握了补习获得的知识,

从而错过了在课堂上进行思维训练和知识构建的黄金时期。因此,农村或经济欠发达地区应该是课外补习重点治理的区域。首先,这些弱势群体难以负担起高成本的高质量补习,政府可以给家庭社会经济地位较低、学业成绩较差的学生提供必要且免费的补习机会。其次,政府可以效仿韩国的“课后教育计划”,资助全职教师使用放学后时间为低收入家庭的学生“补差”。最后,“安慰剂效应”可以成为课外补习治理的一个突破口。政府应该通过积极的舆论宣传,澄清课外补习的真实效度,缓解人们因从众心理带来的压力知觉。

本文使用了 CEPS 两期的数据,使用了倾向值匹配法和双重差分法,有效克服了不可观测变量和可观测变量带来的内生性问题,获得了较为可靠的因果推断效应。但双重差分倾向值得分匹配法仍然存在一些问题,即无法控制随时间变化的异质性问题。虽然本文已经尽可能地控制随时间变化特征的变量,并且尽量减少数据的时间跨度,但仍无法穷尽上述内生性问题。而且这种做法只能估计课外补习的短期效应,而可能忽视由于补习作用的滞后性带来的积极效果。此外,本文仅使用“是否补习”的二分法对课外补习进行操作化测量。实际上,课外补习的时间、频率、质量乃至补习班的性质都可能影响学生成绩和补习效果。最后,受限于数据的可得性,本文对“安慰剂效应”的测量仅仅是从一般意义上的心理情绪入手,并不是具有针对性的学业情绪,这些都需要未来研究进一步的改进与探索。

参考文献

- [1] 中国教育报. 中国教育学会发布调查报告显示:2016年我国中小学课外辅导“吸金”超八千亿[J]. 教育发展研究, 2017(4):63—63.
- [2][4][17] 薛海平, 丁小浩. 中国城镇学生教育补习研究[J]. 教育研究, 2009(1):39—46.
- [3] 薛海平. 课外补习、学习成绩与社会再生产[J]. 教育与经济, 2016,(2):32—43.
- [5] Popa, S. (2008). Education in a hidden marketplace: Monitoring of private tutoring. *Comparative Education*, 44(4), 525—527.
- [6] 刘菁菁. 东亚国家课外补习的政府监管之道——香港大学教育学院比较教育研究中心主任马克·贝磊教授专访[J]. 外国中小学教育, 2014(10):1—6.
- [7] Dang, H. A. (2007). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review*, 26(6), 684—698.
- [8] Mischo, C. & Haag, L. (2002). Expansion and effectiveness of private tutoring. *European Journal of Psychology of Education*, 17(3), 263—273.
- [9] 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.
- [10] Sunderman, G. L. (2006). Do supplemental educational services increase opportunities for minority students? *Phi Delta Kappan*, 88(2), 117—122.
- [11] Byun, S. Y. (2014). *Shadow education and academic success in Republic of Korea*.

- Korean education in changing economic and demographic contexts*. Springer Singapore.
- [12] Fergany, N. (1994). Quality of life indices for Arab countries in an international context. *International Statistical Review*, 62(2), 187—202.
- [13] Thapa, & Amrit. (2013). Does private school competition improve public school performance? The case of Nepal. *International Journal of Educational Development*, 33(4), 358—366.
- [14] 雷万鹏. 高中生教育补习支出:影响因素及政策启示[J]. *教育与经济*, 2005(1): 39—42.
- [15] 李佳丽, 胡咏梅. 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[J]. *教育与经济*, 2017(2):51—61.
- [16] 张羽, 陈东, 刘娟娟. 小学课外补习对初中学业成绩的影响——基于北京市某初中九年追踪数据的实证研究[J]. *教育发展研究*, 2015(2):18—25.
- [18][28] Zhang, & Yu. (2013). Does private tutoring improve students' national college entrance exam performance? A case study from Jinan, China. *Economics of Education Review*, 32, 1—28.
- [19][29] 李佳丽. 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[J]. *基础教育*, 2018(1): 90—98.
- [20] 黄斌, 方超, 汪栋. 教育研究中的因果关系推断——相关方法原理与实例应用[J]. *华东师范大学学报(教育科学版)*, 2017(4):1—14.
- [21] 刘东芝. 小学生课外补习弊大于利[J]. *中国教育学刊*, 2014(11):19—23.
- [22] 吴愈晓, 黄超. 基础教育中的学校阶层分割与学生教育期望[J]. *中国社会科学*, 2017(3):111—134.
- [23] 万海远, 李实. 户籍歧视对城乡收入差距的影响[J]. *经济研究*, 2013(9):43—55.
- [24] 孟天广, 陈昊. 不平等、贫困与农村基层民主——基于全国400个村庄的实证研究[J]. *公共管理学报*, 2014(2):129—138.
- [25] Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4), 605—654.
- [26] Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), 261—294.
- [27] Heckman, J. J., & Vytalil, E. (2001). Policy-relevant treatment effects. *American Economic Review*, 91(2), 107—111.
- [30] 李文道, 孙云晓. 我国男生“学业落后”的现状、成因与思考[J]. *教育研究*, 2012(9): 38—43.
- [31] 李春玲. “男孩危机”“剩女现象”与“女大学生就业难”——教育领域性别比例逆转带来的社会性挑战[J]. *妇女研究论丛*, 2016(2):33—39.

(责任编辑 范皓皓)

The Effectiveness of Private Tutoring: An Estimation Based on China Education Panel Survey

SUN Lunxuan, TANG Jingjing

Page 123

The question of whether private tutoring can effectively improve academic performance has come into focus in recent years. Based on China Education Panel Survey (CEPS) data, this study finds that private tutoring does not have a positive impact on academic achievement of middle school students in China. On the contrary, it has a negative impact to some extent. At the same time, we find that private tutoring significantly reduces the frequency of students' negative emotions such as feeling blue, feeling unhappy, not enjoying life and feeling sad. The above findings suggest that private tutoring has only emotional value rather than educational value for most participants, and we believe that this is mainly due to the collective anxiety caused by educational competition. Further analysis finds that emotional value mainly exists for disadvantaged groups of students, namely, those who are male and those who come from rural areas or non-elite families. At the same time, private tutoring has no academic effect and even has a negative effect on these students. Thus, private tutoring appears to benefit advantaged group and harm disadvantaged groups. From this perspective, private tutoring further enlarges the inequality in academic achievement among different groups and should therefore receive more attention from the Chinese government and families.

Re-analysis of Regional Differences between Enrollment Expansion and Higher Education Enrollment Opportunities

ZHANG Donghai, LI Li

Page 142

This paper first classifies 31 provincial administrative regions into three categories—regions of absolute superiority, relative superiority, and relative inferiority respectively—according to gross enrollment rate and high-quality enrollment rate of higher education. The Wilson score and Theil index for admission opportunity are calculated to reveal the change in regional disparity since the enrollment expansion. Based on statistics and questionnaire, this