

重点班能否提升高中生的学业成绩？

——基于西部某市普通高中的断点回归设计研究

杨 靖，叶晓梅，杜育红

[摘要] 设立重点班是普通高中教育实践中常见的资源配置策略，但由于缺乏实证依据，其公平与效率问题长期备受争议。本文采用我国西部某市五所县域普通高中2017和2018两届学生的追踪数据，运用断点回归方法，从提升学生学业成绩的角度检验设置重点班的有效性。研究发现，尽管重点班学生高一分班和高三时的分数均高于普通班学生，但此期间重点班与普通班学生的成绩差距没有显著拉大。进一步，利用重点班分数线这一外生变量，估计高一分班分数线附近能力类似，但分别进入普通班和重点班的两类学生在高三时的成绩差异。参数估计和非参数估计结果均显示进入重点班没有显著提高学生的学业成绩，在男女、城乡学生间也未发现异质性差异。由此可知，普通高中的重点班并非是优化教育资源配置与提升学生学业成绩的有效策略，更可能是资优生与普通生成绩差距的固化器。未来普通高中班级设置需兼顾公平与效率，促进不同类型学生的增值发展。

[关键词] 重点班；学业成绩；断点回归设计

一、引言

长期以来，普通高中设置的“清北班”“宏志班”“火箭班”等重点班为家长所青睐，是学校吸引优质生源、提高升学率和扩大社会影响力的重要手段。实践中，学校依据学生成绩将资优生分配进重点班，并优先配备经验丰富、具备声望的教师和其他优质教育资源，实施培优教学。“望子成龙，望女成

[收稿日期] 2021-08-18

[基金项目] 教育部哲学社会科学研究重大课题攻关项目“教育与经济发展关系及贡献研究”(15JZD040)。

[作者简介] 杨靖，北京师范大学教育学部教育经济研究所/四川省天全中学，电子邮箱地址：201523010066@mail.bnu.edu.cn；叶晓梅(通讯作者)，北京大学教育学院，电子邮箱地址：iiker@163.com；杜育红，北京师范大学教育学部教育经济研究所，电子邮箱地址：dyh@bnu.edu.cn。

风”的中国式家长对重点班十分偏爱，并笃信重点班能帮助子女提升学业成绩。但这些想法多基于日常经验的判断，并且这种区别化的教育资源配置方式，也因违背教育公平的价值追求遭到批评（Vanderhart, 2006; Betts, 2011）。重点班能否提升高中生的学业成绩，当前仍难以从实证研究中获取参考。

从资源配置理论上讲，将最优的资源集中配置是提升劳动生产率的重要策略。但这种工业化生产的思想忽视了学生作为全人的特征，学校设置重点班的教育资源配置方式还需考虑学生的心理发展特征。Davis(1966)提出的“青蛙池塘”理论认为，学生更倾向于基于团体内而非团体间的同伴能力来评估自身能力。如果学生在精英学校中排名比较靠后，往往会遭受同伴压力并导致学业下滑。但如果其选择在一般学校就读，会因排名优于同伴而增加自信心，进而对其学业产生有利的影响。此外，部分相关实证研究也发现重点学校(班级)对学生学业成绩并不能产生显著的提升作用(Abdulkadiroglu et al., 2014; Dee and Lan, 2015; Shi, 2020)。学校内部资源配置中重点班是否会产生同样的效果，还有待本土化的证据进一步验证。

在教育公平和效率的理念下，科学测算重点班对学生学业发展的真实影响效应，不仅能为家长和学生理性教育决策提供依据，也可以为学校教育资源配置提供循证依据。因此，本文将基于我国西部某市五所县域普通高中重点班和普通班学生的发展状况，依据准随机实验思想，利用断点回归分析方法的来评估重点班这一学校内部资源配置策略对学生学业成绩的影响效应。

二、文献回顾与研究假设

重点班或重点学校的现象并非国内独有，在世界各地的基础教育系统中普遍存在，学界通常在能力分流/分组(Ability Tracking/Grouping)视角下关注。在教育公平和效率理念下，各国研究者尤为关注重点班(校)对学生学业成绩的影响效应。但要得到重点班对学生学业成绩的“净”影响通常面临遗漏变量、样本自选择等问题的挑战，直接采用最小二乘估计得到的结论通常是有偏的。研究者多依据准随机实验的思想，采用工具变量和断点回归的方法进行估计，也有少数学者采用随机实验的方法研究该问题。但未取得一致的结论，主要存在两类观点。

部分学者认为，重点班(校)能对学生的学业发展产生了积极的影响，并

将其归因于学业成绩更高的同伴, 优质和丰富的学习资源和更具挑战性的课程设计, 来自美国和冰岛等地的实证研究均证实了这一结论。如 Hastingsh 和 Weinstein(2007)以北卡罗来纳州夏洛特梅克伦堡学区通过随机排位机制获得重点学校转入许可的 6695 名中小學生为研究对象, 发现转入重点学校就读的学生较之未转入的学生, 学业成绩显著提升了 1 个标准差。Vardardottir (2013)运用冰岛一所高中的五年追踪数据和断点回归估计方法, 也发现进入重点班使学生在年度考试和春季考试的成绩上分别提高 0.47 和 0.32 个标准差。基于国内普通高中的研究也发现了类似的结论, 如 Ding 和 Lehrer (2007)以江苏省 10 所高中的学生为研究对象, 通过准随机实验设计, 采用断点回归估计重点高中对学生的影响, 发现相似能力的学生, 进入重点高中显著提升了学生的高考成绩。王俊和孙志军(2015)利用 F 县两届普通高中学生样本, 采用断点回归设计, 发现进入重点高中对理科学生数学和语文成绩有显著的促进作用。而 Wu 等(2019)以湖南省某县两所重点高中三届共计四千多名学生为研究对象, 通过断点回归估计也发现重点班能显著提高学生的学业成绩。

然而, 也有部分研究发现重点班(校)并不能对学生的学业发展产生显著的积极影响。这可能是因为在重点班(校)的优质同伴群体中, 学业成绩排名靠后, 同伴往往引发其学习上的压力并使其士气受损, 进而对学生的学业发展产生不利影响, 即“歧视性对比”(Marsh et al., 2000)。同时, 重点班(校)的课堂教学往往强调更高难度的学习内容和更快的学习节奏, 易对班(校)内成绩排名靠后的学生造成不利的影响(Duflo et al., 2011; Bui et al., 2014)。来自国外的部分实证研究也证实了进入重点班级(校)并不能显著提升学生的学业成绩。如 Cullen 等(2005)利用芝加哥地区公立学校的 60000 多名高中生样本, 利用随机排位的入学机制构建工具变量, 发现能力近似的学生进入精英学校和普通学校后, 学业成绩并未呈现出显著差异。Duflo 等人(2011)利用肯尼亚 121 所小学学生的追踪数据, 采用断点回归方法估计重点班分数线附近进入和未进入重点班学生的学业成绩差异, 结果发现重点班并没有显著提升学生的学业成绩。Abdulkadiroglu 等人(2014, 2017)分别基于美国波士顿、纽约和芝加哥精英考试学校(Exam School)申请者的研究, 同样发现精英考试学校没有对学生学业成绩产生显著的积极影响。Shi(2020)基于美国北卡罗来纳州 13 个学区特许学校(Chart School)和公立学校的随机试验研究也得到了相似的结果。来自我国部分地区的实证研究也证实了这一结论, 如 Dee 和 Lan(2015)使用内蒙古某市多届高中学生的追踪数据, 利用重点高中录取分数线这一外生变量, 采用模糊断点回归设计, 发现

就读于重点高中没有显著提升学生的高考成绩。Zhang(2016)基于长江流域某发达省会城市八所重点初中三届,共计13768名申请学生的追踪研究,也发现学生进入重点中学并未对学业成绩产生显著的积极影响。

综上,重点班对学生学业成绩的影响效应,并没有取得一致的结论。有学者将其归因于测量和样本特质的问题(Lee and Lemieux, 2010),这促使我们深入到具体的教育背景下探究重点班对学生学业成绩的影响效应。鉴于当前研究缺乏关注中国西部欠发达地区县域普通高中重点班对学生学业发展的影响,本文以我国西部某省Y市五所县域公立高中为样本,探究重点班对学生学业成绩的影响效应,期望基于断点回归方法的估计结果,为西部地区普通高中学校内部的班级配置和家长教育选择提供参考依据。基于已有研究中存在相关的结论,提出以下研究假设:

H_0 : 重点班不能显著提升学生的学业成绩。

H_1 : 重点班能显著提升学生的学业成绩。

三、数据和研究方法

(一)数据介绍

本文的数据来源于我国西部某市2017届和2018届高中应届理科生追踪信息库。其中包含了学生的性别、年龄、户籍、所在学校、全市统考的分班考试成绩和高三二诊考试的成绩,也包含了学生在重点班和普通班的分班信息。该市共有公立普通高中12所,其中一所直属市教育局,其余11所分属8个区县教育局。鉴于本研究重点关注县域普通高中理科重点班的影响效应,故排除市辖区学校3所,以艺体生培养为主的学校2所(该省高考艺体生文化课分数线以文科高考分数为基准划定,不在本研究范围内)。此外,各校会依据教育主管部门要求和生源情况,给各班下达不同的高考上线指标;受此影响,教师在教学中可能会更加关注上线指标覆盖范围内学生的学习需求。为克服这种“教师自选择”可能对估计结果造成的干扰,研究排除了2014年和2015年高考理科本科上线人数少于重点班规模的学校3所,最终选取5所县域普通高中作为研究样本。

样本高中在学生高一结束时依据学生填报志愿对学生进行文理科分流,学校会依据学生填报的志愿情况设定文理科重点班和普通班的数量和规模,部分学校还会设定艺体班(在文科班中分流)的数量和规模,这一工作通常在学生高一下学期全市期末统考前完成。期末统考结束后,学校对学生在期末统考中文理科相应高考科目的总分进行排名,并依据事前确定的重点班规模

和相应排名学生的成绩划定重点班分数线,其中理科重点班的分数线依据语文、数学、英语和理科综合的总分划定。当学生总分高于重点分数线则进入重点班,低于分数线则进入普通班。同时,同层次班级间则依学生成绩排名,按“Z”字型方式^①进行分配。重点班分数线的设置使得以分班成绩为条件,学生进入重点班就读的概率分布在分数线处不连续,存在跳跃,为本研究使用断点回归设计提供了契机。

本研究的研究对象为5所学校2017届和2018届的全体理科生,共计3397人,其中2017届1669人,2018届1728人。各学校2017届和2018届的理科生规模、分班成绩和重点班分数线如表1所示。为了便于比较,我们在全市理科生层面对分数进行了标准化处理。数据显示各校间理科生的生源状况虽有较大差异,但由于各校内部设立重点班的数量和规模(C校两个班,其余一个班)以及学生与学生成绩的异质性程度不同,各校划定的重点班分数线较为接近,都在0.64至0.96个标准差之间。

表1 各学校、各届理科生分班成绩的描述性统计和重点班分数线

学校	年级	平均分	个案数	标准差	重点班分数线
A	2017	0.683	364	0.532	0.96
	2018	0.614	442	0.686	0.93
B	2017	-0.151	257	1.036	0.87
	2018	-0.170	260	1.025	0.89
C	2017	-0.105	388	0.803	0.64
	2018	-0.055	361	0.904	0.67
D	2017	-0.337	317	1.024	0.94
	2018	-0.162	313	0.923	0.91
E	2017	0.272	343	0.740	0.95
	2018	0.244	352	0.778	0.93
总计	2017	0.093	1669	0.907	
	2018	0.140	1728	0.908	

注:1. 平均分、标准差和重点班分数线是全市理科生总分标准中心化后的成绩;
2. 考试成绩为文理分科前一次全市统考成绩理科高考科目的总分。

样本结构是本研究是否适用断点回归的另一基本条件。由于实际的分班过程中,各学校对上述分班规则的执行程度有差异,可能存在部分学生

^① 如:第1名进入1班,第2名进入2班,第3名进入3班,则第4名进入3班,第5名进入2班,第六名进入1班,以此类推。

会因为先前高中学习中已经适应了原班级的各科教师，或是存在“鸡头凤尾”的博弈，即使成绩高于分数线也不愿意进入重点班级。还有部分学生会利用家长的社会资源和相关措施，以低于重点班分数线的成绩进入重点班。如果学生总成绩超过分数线未进入重点班就读以及未达分数线却进入重点班就读的学生比例过大，将使分数线两侧进入重点班的学生比例分布没有明显跳跃，断点回归设计失效，因此需对样本结构进行考察。由表2可以看出两届学生中，低于重点班分数线但在重点班就读的比例皆低于30%，而高于重点班分数线但未在重点班就读的学生比例低于10%，符合研究需要。

表2 样本构成

年级	样本学生数	重点班人数	低于分数线进入重点班	高于重点班分数线未进入
2017届	1669	268(16.05%)	65(24.25%)	119(8.42%)
2018届	1728	351(20.31%)	73(20.80%)	111(8.06%)

注：括号外为人数，括号内为百分比。

表3是重点班和非重点班学生在周课时时间、班级规模和性别等方面的描述性统计。在周课时时间分布上，2017届平均在61小时以上，而2018届平均在60小时以上。班级规模方面，2017届重点班规模接近46人，2018届约50人，两届非重点班规模都超过55人。性别方面，重点班的女生比例高于普通班，重点班女生比例为51.3%和57.1%，而在普通班则为46.7%和45%。两届学生在各控制变量上的差异很小，一定程度上反映出学校教育生产环境在两届学生中并未发生显著变化，

表3 重点班与非重点班控制变量的描述统计

届别	班级类型	周课时时间	班级规模	性别
2017	普通班	61.252	57.866	0.467
	N=1401	(0.080)	(0.142)	(0.013)
	重点班	61.060	45.515	0.571
	N=268	(0.172)	(0.327)	(0.033)
2018	普通班	60.415	55.447	0.450
	N=1377	(0.038)	(0.291)	(0.013)
	重点班	60.123	49.818	0.513
	N=351	(0.068)	(0.185)	(0.027)

注：括号内为标准误。

(二) 研究设计

在识别重点班对学生学业成绩影响效应中,最常见的问题是遗漏变量和样本自选择对估计结果造成的偏误。要得到重点班对学校学业成绩的“净”影响,需使学生在重点班和普通班间随机分配,但这在实践中通常难以实现。研究人员通常采用的识别策略是寻找外生变量,并应用工具变量和断点回归技术来构建随机实验环境,也即让在分数线两侧某个范围内影响学生学业成绩的可观察和不可观察的特征在两类样本间相似,这样就使是否进入重点班成为造成学业成绩差异的原因。在本研究中,各学校的重点分数线是根据考试前设定的重点班规模和数量,以及全校学生在高一期末的全市统考中的表现决定的,学校和学生并不能操纵它来决定某个学生是否进入重点班,可以将其视为随机实验实施前后施加给条件样本学生的一个外生变化,并利用分数线两侧较小的范围内不可观测因素类似的学生构建准随机实验场景。

原则上,学生学业成绩只要达到分数线就能进入重点班,反之则进入普通班。但由表2的样本结构可知,在实际的分班过程中,样本中存在学生总成绩超过分数线未进入重点班就读以及未达分数线却进入重点班就读的现象,这样就使以分数线为断点两侧的不连续性跳跃不是1(进入重点班)和0(未进入重点班)。由此,研究采用模糊断点的识别策略进行估计,利用分数线作为工具变量,估计录取规则遵从者间的学业成绩差异。同时, Lee 和 Lemieux(2010)指出,参数和非参数估计是互补的,提供了一个相互加强的规范检查。研究分别通过非参数估计和参数估计来识别重点班对学生学业成绩的影响。鉴于重点班分数线在每个学校、每届间有差异,研究在估计模型中加入届一校的固定效应,控制在校际间/届别间不随时间变化但对学业成绩造成影响的因素。

1. 断点回归思路

模糊断点回归设计通过估计是否进入重点班对学业成绩影响的局部平均处理效应(LATE)来探寻这对变量的因果关系,等价于公式(1):

$$ATE_{FRD} = E(y_{1i} - y_{0i}) \mid \text{个体是遵从者且 } A = \bar{A} \quad (1)$$

遵从者即高于或等于重点班分数线的学生进入到重点班学习,而低于分数线的学生进入普通班级学习。由于个体在考试前无法控制分数线 c 值,满足给定驱动变量 A 时,个体进入实验组的决定独立于参与所得到收益的假设,可以通过公式(2)得到平均处理效应 ρ :

$$\rho = \frac{\lim_{h \rightarrow 0} E(y_i | c < x_i < c+h) - \lim_{h \rightarrow 0} E(y_i | c-h < x_i < c)}{\lim_{h \rightarrow 0} E(D_i | c < x_i < c+h) - \lim_{h \rightarrow 0} E(D_i | c-h < x_i < c)} \quad (2)$$

重点班对学生学业成绩的影响为 ρ ，即局部平均处理效应(Local average treatment effect, LATE)。驱动变量 x_i 为学生 i 分班前一次全市统考中心标准化后的成绩；处理变量 D 为是否进入重点班，进入为 1，否为 0； y_i 为学生 i 在分班后全市统考中的学业成绩。H 为带宽，基于上述局部连续性假设，在进入重点班的概率上有跳跃，设定带宽为 $h > 0$ ，在分数线 $[c-h, c+h]$ 的区域内就可对其进行估算。

由此本文的驱动变量的计算方式如公式(3)， c_{js} 代表 s 学校 j 届重点班的分数线。

$$A_{-ijs} = x_{ijs} - c_{js} \quad (3)$$

2. 参数估计

参数估计实质上是构建工具变量的多项式回归。分数线是各学校依据该校学生统考成绩的校内排名和预先设置的重点班名额划定的。对于学生来说在考试前并不知道且无法决定考到多少分数能够进入重点班级；对于学校来说，学校即使对某类学生有一定偏向，也无法在考前对分数线进行控制。因此，分数线是一个完全外生的变量，可构建工具变量，通过两阶段最小二乘法(2SLS)进行估计就可以解决由于学生个体特征、选择偏好等内生性问题对估计带来的偏误。我们只需要将样本限制在分数上下一个较小的区域，就可以认为这部分学生不可观测的因素有较强的相似性，从而得出重点班对学生学业成绩的净影响，其基本模型如下：

第一阶段回归：

$$D_{ijs} = \gamma_0 + \gamma_1 T_{ijs} + (1 - T_{ijs}) f_0(A_{-ijs}) + T_{ijs} f_1(A_{-ijs}) + \gamma_4 X_{ijs} + \delta + \mu_{ijs} \quad (4)$$

第二阶段回归：

$$Y_{ijs} = \alpha_0 + \rho_1 D_{ijs} + (1 - T_{ijs}) f_0(A_{-ijs}) + T_{ijs} f_1(A_{-ijs}) + \alpha_4 X_{ijs} + \delta + \mu_{ijs} \quad (5)$$

其中， T_{ijs} 表示 s 校 j 届 i 学生分班成绩是否超过录取分数线的二分变量，是内生变量 D 的工具变量，即学生是否进入重点班。为最大限度避免遗漏变量对估计结果的影响，同时提高驱动变量和结果变量关于学生成绩在重点班分数线附近的拟合程度，本研究引入中心化处理后统考成绩标准分的多项式 $f(A_{-ijs})$ ， f 下标 1 表示进入重点班的样本，下标 0 表示未进入重点班的样本。为允许上重点班的概率在断点两侧出现斜率的变化，将工具变量 T

和驱动变量 $f(A_{-ijs})$ 的交互项作为控制变量包括在式(4)中, 通过一对多项式进行控制, 定义如式(6), X_{ijs} 为其他控制变量, δ 为届一校的固定效应。

$$f(A_{-ijs}) = \pi_{k1} A_{-ijs} + \pi_{k2} A_{-ijs}^2 + \cdots + \pi_{kn} A_{-ijs}^n \quad k \in \{0, 1\} \quad (6)$$

3. 非参数估计

非参数估计等价于局部线性回归, 与参数估计的不同主要表现在: 首先, 带宽窗口 h 要小于给定带宽下的多项式回归; 其次, 非参数估计使用的是篷形的边缘核 (Tent-shaped edge kernel), 而不是参数估计中默认的一致核 (Uniform kernel); 再次, 非参数模型只控制驱动变量的线性函数, 忽略高阶项。本研究定义的非参数估计模型如式(7):

第一阶段回归:

$$D_{ijs} = \beta_0 + \beta_1 T_{ijs} + \gamma_0 (1 - T_{ijs}) A_{-ijs} + \gamma_1 T_{ijs} A_{-ijs} + \beta_4 Z_i + \delta + \mu_{ijs} \quad (7)$$

第二阶段回归:

$$Y_{ijs} = a_0 + \rho_1 D_{ijs} + \gamma_0 (1 - T_{ijs}) A_{-ijs} + \gamma_1 T_{ijs} A_{-ijs} + \alpha_4 Z_{ijs} + \delta + \epsilon_{ijs} \quad (8)$$

非参数模型控制驱动变量的一个线性函数, 按照 Imbers 和 Kalyanaraman(2012)的方法计算最佳带宽, 并使用篷形边缘核, 中心围绕表示为 $K(A_{-ijs})$:

$$K(A_{-ijs}) = I\left\{\frac{A_{-ijs}}{h} \leq 0\right\} \left(1 - \left|\frac{A_{-ijs}}{h}\right|\right) \quad (9)$$

另外, 使用模糊断点回归进行估计时还需要通过图像和不同带宽值估计的报告, 关注断点估计结果的稳健性和对带宽变化的敏感性。同时还要满足驱动变量和控制变量在断点处的连续性假设, 即除了处理效应(重点班)其他特征在两类学生中无系统性差异。

四、实证分析

(一) 重点班和非重点班学生的差异

对学生分班考试成绩和高三成绩进行全市层面的标准化处理后进行描述统计。根据表 4 的分析结果, 发现重点班学生的语文、数学、英语和理科综合平均成绩均显著高于普通班学生约 1 到 1.4 个标准差, 这一结果在 2017 届和 2018 届学生群体中均呈现出一致的趋势。如前所述, 造成这种差异的原因并不一定是由学生是否进入重点班学习造成, 它还混淆了学生家庭背景、学习基础和学习能力等多个因素, 因此通过描述性统计和 T 检验并不能识别重点班对学生学业成绩的影响。

表4 重点班与非重点班学生的学业成绩差异

届	考试	班级类型	语文	数学	英语	综合	总分
2017	高一分班	普通班	-0.236	-0.059	-0.236	-0.067	-0.100
		n=1401	(0.024)	(0.025)	(0.024)	(0.023)	(0.023)
		重点班	1.168	0.94	1.168	1.104	1.216
	高三二诊	n=268	(0.033)	(0.036)	(0.033)	(0.038)	(0.032)
		二者差异	-1.403***	-0.999***	-1.403***	-1.172***	-1.316***
		普通班	-0.189	-0.059	-0.189	-0.121	-0.115
2018	高一分班	n=1401	(0.024)	(0.025)	(0.024)	(0.023)	(0.023)
		重点班	1.164	1.022	1.164	1.138	1.228
		n=268	(0.043)	(0.050)	(0.043)	(0.042)	(0.038)
	高三二诊	二者差异	-1.353***	-1.082***	-1.353***	-1.259***	-1.343***
		普通班	-0.168	-0.146	-0.168	-0.105	-0.135
		n=1377	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)
2018	高一分班	重点班	0.997	1.174	0.997	1.191	1.191
		n=351	(0.03)	(0.039)	(0.030)	(0.032)	(0.032)
		二者差异	-1.165***	-1.320***	-1.165***	-1.296***	-1.326***
	高三二诊	普通班	-0.144	-0.121	-0.144	-0.156	-0.143
		n=1377	(0.024)	(0.024)	(0.024)	(0.024)	(0.022)
		重点班	1.034	1.268	1.034	1.162	1.248
高三二诊	n=351	(0.036)	(0.044)	(0.036)	(0.041)	(0.036)	
	二者差异	-1.178***	-1.388***	-1.178***	-1.319***	-1.391***	

注：1. * 为 0.1 水平上统计显著，** 为 0.05 水平上统计显著，*** 为 0.01 水平上统计显著；2. 括号内为标准误。

(二) 基于断点回归原理估计重点班对学生学业成绩的影响

1. 断点识别

应用断点回归设计的关键条件是驱动变量的期望值在断点两侧必须存在跳跃。图 1 和图 2 分别为两届学生进入重点班比重和分班总成绩标准分以及语文、数学、英语和理科综合成绩标准分的关系图，其中，横轴表示以重点班分数线做中心化处理后的总分及各科成绩的标准分，纵轴为重点班学生所占的比重。在重点班分数线处，进入重点班学生所占的比重从不到 20% 骤然上升到 45%，各学科也出现了类似比重值的上升，出现了十分明显的跳跃，说明在重点班分数线处存在断点。

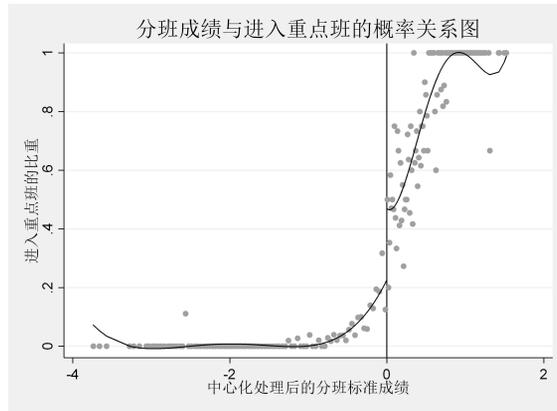


图 1 总分断点示意图

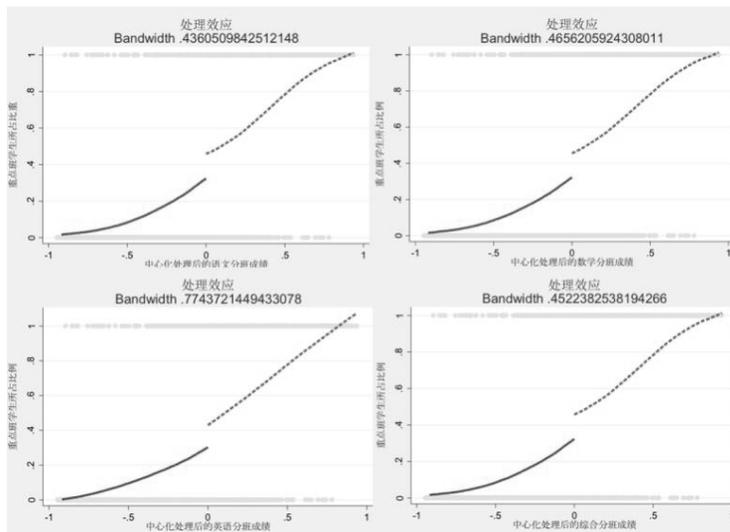


图 2 各科分数断点示意图

2. 带宽和多项式次数的选取

局部线性回归的关键是估计过程中最优带宽的选择, 研究分别采用 Calonico 等人(2014b)(CCT 方式)、Imbens 和 Kalyanaraman(2012)(IK 方式)以及 Ludwig 和 Miller(2007)(CV 方式)共三种方式对最优带宽进行选择, 在全样本下以总成绩标准估计最优带宽 h , CCT 的估计值为 0.945, IK 为 1.176, CV 为 1.64。在进行局部多项式回归时, 按照 Hahn 等人(2001)的经验原则, 涵盖的断点附近的样本量应该在总样本量的 2.6%—35.2%之间, 对应到本文中, 包含的样本量应在 88—1196 人。依据此原则, 通过非参数估

计(CCT)得到总成绩分布的最优带宽值为 0.945 个标准差,其包含的样本量在这一区间之内,由此在参数估计部分将在 CCT 最优带宽值下进行估计。此外,还将选取 50%最优带宽(0.473 个标准差,样本数 1076)和 25%最优带宽(0.237 个标准差,样本数 568)分别进行参数估计以检验估计结果的稳健性。另一方面,依据赤池信息准则(AIC 值最小)选取各带宽下的多项式次数。但 Gelman 和 Imbens(2019)指出,断点回归估计中的高阶多项式函数通常很不可取,易导致误导性的置信区间。由此,进入模型的多项式次数不超过 3 次。表 5、表 6 中报告了在各带宽下对总分和各科成绩进行参数估计时的多项式次数的选取,最优带宽下各科和总分多项式次数为 3 次,50%最优带宽下为 2 次,25%最优带宽下为 1 次。

3. 估计结果

根据研究设计,依次采用非参数和参数估计的方法来探究重点班对学生学业成绩的影响。图 3 报告了进入重点班和未进入重点班学生在高三二诊考试中的差异,可以看出无论是总分,还是各学科的成绩,在断点处都未出现明显的跳跃。

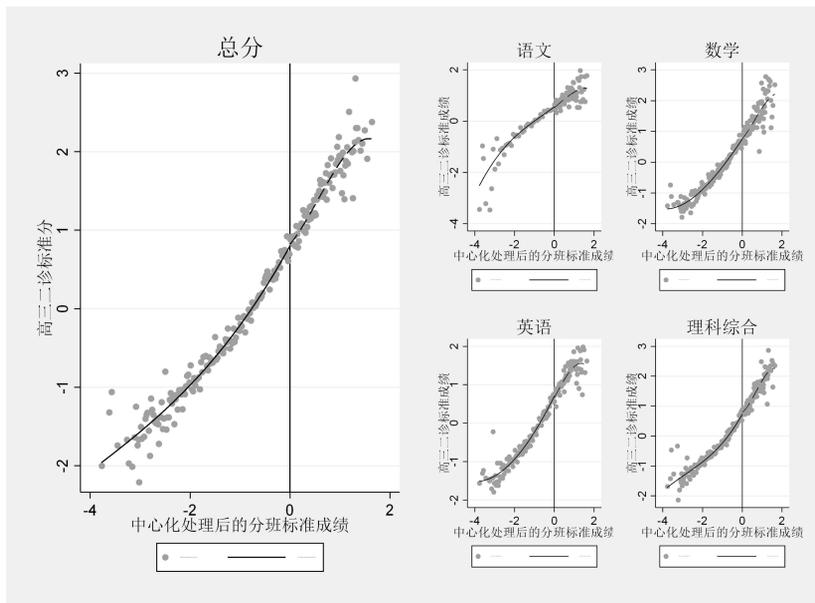


图 3 重点班和非重点班的学生成绩差异

同时,表 5 报告了在最优带宽内重点班对学生学业成绩的影响。左边列为非参数(核估计)估计结果,右边列为引入工具变量的两阶段最小二乘(2SLS)参数估计结果。参数估计和非参数估计的结果差异很小,估计结果是

稳健的,这进一步印证了图 3 的事实,重点班这种政策效应对学生统考总分和各科成绩的影响都没有达到显著水平,不能拒绝政策效应 $\rho=0$ 的假设。这与 Cullen 等人(2005)、Duflo 等人(2011)以及 Dee 和 Lan(2015)的研究结果一致——相似学业背景的择校生和常规入校学生间的学业成绩没有显著差异,并认为优势学校学生与普通学校学生间出现的学业成绩差异主要在于学习动机、学习能力和父母关注等内生因素的影响。

在影响幅度方面,学科间表现出不同的幅度和方向。在数学和综合成绩方面表现出积极的影响,但影响幅度微乎其微,非参数估计的效应为 0.059 和 0.019 个标准差,参数估计则分别为 0.062 和 0.038 个标准差,换算成原始分两科成绩的影响幅度都不足两分。而在总分、语文及英语学科则表现出消极影响,影响幅度也极小,非参数和参数估计发现对总分的影响都是约 0.05 个标准差,换算成原始分,影响幅度大约 5 分;对语文学科的非参数和参数估计结果分别是 -0.152 和 -0.220 个标准差,换算成原始分,影响幅度大约两分;对英语学科的非参数和参数估计结果分别是 -0.171 和 -0.202 个标准差,换算成原始分,影响幅度大约 4 分,但这些影响并不具有统计学意义。

表 5 最优带宽下参数和非参数估计重点班对学生学业成绩的影响

	总分		语文		数学		英语		综合	
	非参数	参数	非参数	参数	非参数	参数	非参数	参数	非参数	参数
处理效应	-0.051 (0.116)	-0.049 (0.203)	-0.152 (0.139)	-0.220 (0.341)	0.059 (0.175)	0.062 (0.302)	-0.171 (0.181)	-0.202 (0.287)	0.019 (0.165)	0.038 (0.257)
控制变量	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R 方	0.602		0.167		0.372		0.391		0.5	
最优带宽	0.945		0.436		0.466		0.774		0.452	
置信区间	(-0.389, 0.508)		(-0.607, 1.169)		(-0.429, 1.091)		(-0.577, 0.212)		(-0.411, 0.829)	
多项式次	3		3		3		3		3	
样本数	1907		1907		1907		1907		1907	

注: 1. * 为 0.1 水平上统计显著,** 为 0.05 水平上统计显著,*** 为 0.01 水平上统计显著; 2. 括号内为标准误。

依据经验,将总分接近两个标准差,换算成原始分最大差距约有 190 分的学生看成是一个学业能力相似的群体非常牵强。考虑到在参数估计中带宽样本量的要求,以及已有研究大多采取参数估计方式,因此研究分别在 50% 和 25% 最优带宽下仅以参数估计的方式对重点班政策效应进行稳健性检验。

表6中报告了在50%和25%最优带宽下参数估计的结果。在这两个带宽下的估计结果与最优带宽估计结果相似,对总分和各学科的影响方向并未发生改变;虽随着带宽的减小,影响幅度有所增加,但均未表现出显著性,不能拒绝重点班政策效应 $\rho=0$ 的假设,也即进入重点班并未对学生学业成绩起到显著促进作用。

表6 参数估计一半和四分之一最优带宽重点班对学生学业成绩的影响

最优带宽	总分		语文		数学		英语		综合	
	0.5%	0.25%	0.5%	0.25%	0.5%	0.25%	0.5%	0.25%	0.5%	0.25%
处理效应	-0.090 (0.454)	-0.155 (0.525)	-0.099 (0.833)	-1.486 (1.737)	0.635 (0.765)	1.424 (1.562)	-1.085 (0.824)	-1.422 (0.302)	0.476 (0.608)	0.409 (1.026)
控制变量	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R方	0.342	0.100	0.100	0.031	0.051	0.011	0.015	0.004	0.198	0.068
多项式次	2	1	2	1	2	1	2	1	2	3
样本数	1076	568	1076	568	1076	568	1076	568	1076	568

注:1.*为0.1水平上统计显著,**为0.05水平上统计显著,***为0.01水平上统计显著;2.括号内为标准误。

4. 稳健性检验

第一,控制变量的连续性检验。控制变量在断点处不受重点班这种政策效应的影响,即在断点处是连续的、没有跳跃是断点回归的重要假设之一。以往的研究一般通过作图和回归分析的方式进行检验,本文采用回归分析的方式对控制变量周课时时间、班级规模和性别进行稳健性检验,结果见表7。通过断点处的密度函数估计,三个控制变量在断点两侧的密度均不存在显著差异,不能拒绝它们的密度函数在断点处是连续的假设。

表7 控制变量断点处的概率密度函数

总分	周课时时间	班级规模	性别
系数	0.327 (2.416)	-9.490 (7.368)	-0.850 (0.643)

注:1.*为0.1水平上统计显著,**为0.05水平上统计显著,***为0.01水平上统计显著;2.括号内为标准误。

第二,驱动变量的连续性检验。驱动变量在断点处的连续性也是断点回归的重要假设之一。本文排除掉分班规则的不服从者后(即高于重点班分数线不在重点班就读或低于重点班分数线在重点班就读的学生),采用McCrary(2008)的方法检验驱动变量的密度函数在断点处(分数线)是否连续。估计结果表明, $\hat{\theta}$ 为0.166,标准误为0.166,没有显著性。标准化分班成绩

在断点处的密度函数如图 4 所示, 密度函数估计值的置信区间在断点(分数线)两侧绝大部分是重叠的, 不能拒绝驱动变量在断点处是连续的假设。

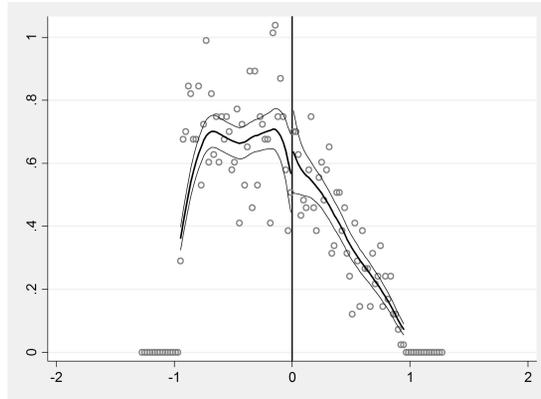


图 4 标准化分班成绩密度函数在断点处的连续性

第三, 人为断点值的处理效应检验。断点回归的基本假设是在没有处理干预的情况下, 对照组(未进入重点班)和处理组(进入重点班)的可估回归函数是连续的。研究者通常需在断点值以外重新设定断点值(也叫作安慰剂断点或者伪断点), 检验在安慰剂(伪)断点处是否存在显著的处理效应, 也即可估回归函数在安慰剂(伪)断点处是否存在跳跃。本文首先将样本限制在进入重点班的学生, 以驱动变量 0.45、0.25 和 0.1 为安慰剂(伪)断点进行检验。然后将样本限制在未进入重点班的学生, 以驱动变量 -0.45、-0.25 和 -0.1 为安慰剂(伪)断点分别进行检验, 结果如表 8。六个安慰剂断点的估计结果皆显示不存在显著性, 不能拒绝在真实断点值外可估回归函数的连续性假设。

表 8 安慰剂断点的估计结果

安慰剂断点值	-0.45	-0.25	-0.1	0.1	0.25	0.45
处理效应	-0.076	-0.060	-0.142	0.150	-0.318	-0.235
	(0.102)	(0.097)	(0.139)	(0.480)	(0.426)	(0.392)
控制变量	yes	yes	yes	yes	yes	yes
最优带宽	0.153	0.13	0.091	0.07	0.064	0.161
样本数	199	180	153	77	75	136
	209	187	97	78	60	100

注: 1. * 为 0.1 水平上统计显著, ** 为 0.05 水平上统计显著, *** 为 0.01 水平上统计显著; 2. 括号内为标准误。

第四, 还需要检验断点处的跳跃是否是因为非常靠近断点的样本特殊性所致, 即断点附近观测值对估计结果的敏感性。检验此假设通常采用“甜甜圈”法(donut hole), 即排除断点附近一定范围内的样本后对模型进行估计, 观察断点附近观测值是否对估计结果造成影响, 而由断点处往外延伸的带宽值被叫做甜甜圈半径。研究分别以驱动变量 0.02、0.04、0.06 和 0.08 为半径分别进行估计, 结果如表 9。四个半径下的处理效应估计结果皆不显著, 说明并不存在断点附近特殊样本对估计结果的系统性影响。

表 9 “甜甜圈”估计结果

半径	0.02	0.04	0.06	0.08
处理效应	0.019 (0.067)	0.041 (0.068)	0.029 (0.068)	0.043 (0.072)
控制变量	yes	yes	yes	yes
最优带宽	0.94	0.96	0.97	1.09
样本数	1863	1891	1884	1931

注: 1. * 为 0.1 水平上统计显著,** 为 0.05 水平上统计显著,*** 为 0.01 水平上统计显著; 2. 括号内为标准误。

最后, 从表 1 可以看出 C 校的重点班分数线要低于其他四校的重点班分数线约 0.3 个标准差。全样本估计中, C 校的学生样本与其他学校的学生样本在学习基础上存在差异, 有可能对估计结果造成影响。由此, 需将 C 校样本剔除后观察估计结果是否稳健。研究通过非参数回归, 在 CCT 最优带宽下进行估计, 结果如表 10。从总分和各科目的估计结果可以看出, 虽然估计值的大小发生了改变, 但影响方向和显著性皆未发生改变, C 校的学生样本并未对总体的估计结果造成显著影响。综上, 本文的断点设计是合适且有效的。

表 10 去掉 C 校后的估计结果

	总分	语文	数学	英语	综合
处理效应	-0.164 (0.246)	-0.112 (0.073)	0.074 (0.318)	-0.048 (0.492)	0.359 (0.368)
控制变量	yes	yes	yes	yes	yes
最优带宽	0.76	0.64	0.81	0.56	0.75
置信区间	(-0.648, 0.319)	(-0.254, 0.030)	(-0.823, , 0.674)	(-1.013, 0.916)	(-0.363, 1.082)
样本数	1485	1165	1436	1042	1337

注: 1. * 为 0.1 水平上统计显著,** 为 0.05 水平上统计显著,*** 为 0.01 水平上统计显著; 2. 括号内为标准误。

5. 异质性检验

利用全体样本进行估计得到的无显著性的局部平均处理效应, 其普适性值得商榷。它有可能是亚群体间正反作用相互抵消所致, 也有可能是在个别亚群体中表现出显著性, 但在全样本估计中被稀释所致。由此有必要检验重点班政策效应在不同特征群体中的异质性。研究分别将不同性别和不同户籍的学生样本代入模型进行异质性检验, 估计结果如表 11。从性别上看, 重点班政策效应在总分和各科目上皆没有表现出显著性, 在男生和女生样本中均不能拒绝政策效应 $\rho=0$ 的假设。从户籍类型来看, 重点班政策效应在总分和各科目间同样未表现出显著性, 不能拒绝在城市学生和农村学生中政策效应 $\rho=0$ 的假设。因此, 无论男生还是女生, 城市学生还是农村学生, 进入重点班都没有显著影响他们的学业成绩。

表 11 异质性检验结果

	总分		语文		数学		英语		综合	
	男	女	男	女	男	女	男	女	男	女
处理效应	-0.087	-0.042	-0.256	0.042	0.538	-0.108	0.082	0.220	0.033	0.082
	(0.103)	(0.128)	(0.232)	(0.128)	(0.411)	(0.195)	(0.235)	(0.178)	(0.304)	(0.159)
控制变量	yes									
最优带宽	0.54	0.63	0.39	0.44	0.27	0.4	0.35	0.55	0.38	0.39
R 方	0.625	0.638	0.332	0.360	0.631	0.675	0.656	0.649	0.749	0.748
多项式次	1	1	1	1	1	1	1	1	1	1
样本数	631	720	472	506	346	468	434	598	466	465
	非农	农村								
处理效应	0.141	-0.389	0.162	-0.336	0.306	0.050	-0.137	-0.196	0.207	0.238
	(0.196)	(0.982)	(0.224)	(0.496)	(0.444)	(0.178)	(0.281)	(0.485)	(0.311)	(0.505)
控制变量	yes									
最优带宽	0.36	0.46	0.47	0.42	0.27	0.41	0.51	0.28	0.49	0.37
R 方	0.757	0.769	0.337	0.138	0.498	0.597	0.514	0.477	0.686	0.673
多项式次	1	1	1	1	1	1	1	1	1	1
样本数	396	483	496	470	331	455	602	355	511	467

注: 1. * 为 0.1 水平上统计显著, ** 为 0.05 水平上统计显著, *** 为 0.01 水平上统计显著; 2. 括号内为标准误。

五、结论与建议

在基础教育阶段, 人们普遍观察到重点班学生学业成绩高于普通班, 但

学生学业成绩是个人前期学习基础、学习能力、学习动机、家庭背景和学校投入等多种教育生产要素投入累积的产出结果，只能反映学生当下的学习状况，并不能直接体现重点班这种学校资源配置方式对学生学业成绩的影响作用。探究重点班对学生学业成绩的实际影响作用，不能将学业成绩水平和学业成绩增值混为一谈。本文以西部某市5所县域普通高中2017届和2018届的理科学生为研究对象，采用准随机实验的研究设计，探究了重点班对学生学业成绩的增值影响。

描述性统计结果显示，重点班和普通班学生虽在分班和高三时的成绩上存在较大的差异，但二者在两次考试中的成绩差距并未发生显著变化，说明重点班和普通班学生在此期间，尽管在学校教育资源配置中存在差异，但重点班对学生学业成绩的增值作用有限。为进一步获得重点班对学生学业成绩增值的因果效应，研究利用重点班分数线这一外生变量，建构准随机实验设计，通过模糊断点回归识别重点班对高中生高三时期各科成绩和总分的影响效应。结果发现，重点班对总分和各学科成绩间并未出现显著的促进作用。稳健性检验结果显示，驱动变量和控制变量满足在断点处的连续性假设，并且估计结果未受到不同带宽取值和不同样本构成的影响，在总分和各学科间均未发现重点班对学生学业成绩产生了显著的正向影响，重点班并未对学生学业成绩起到显著的提升作用，研究结果具有较强的稳健性。异质性检验结果表明，重点班同样没有对不同性别和不同户籍类型的学生学业成绩产生显著的正向影响。

基于以上结论，本文得出以下启示：第一，普通高中集中优势资源配置重点班，并不是提高学校高考升学率的有效手段，特别在提高一本升学率上。^①虽然并不能据此提出取消设置重点班的建议，但在教育公平和优质均衡的视野下，普通高中应该在各类型班级间均衡配置资源。第二，单纯使用最终成绩评价重点班与普通班学生成绩的差异并不能鉴别出重点班班级设置方式对学生学业成绩的促进作用，无法为教育教学改进提供可靠的证据和指导。普通高中需提高自身管理水平，完善学校教育质量评价体系，建立以循证为基础的决策机制，探索增值评价在具体学校背景下的实践方式，有效运用评价信息并提高资源配置效率，为教师专业成长提供帮助，提高教师教研

^① 样本市在每次考试中会依据往年高考上线情况划定高考一本与二本的参考分数线，研究中使用的考试成绩的一本预测线在1.2个标准差，断点回归中包含的样本大多为“一本”“踏线生”。

能力, 促进学生学业发展, 激发学校办学活力。第三, 积极落实《关于新时代推进普通高中育人方式改革的指导意见》, 尊重学生差异化的发展需要, 凸显学生在教育公平视域中的主体地位, 以特色取代以学业成绩分班的方式。

[参考文献]

- 王俊、孙志军, 2015: 《重点高中能否提高学生的学业成绩: 基于 F 县普通高中的—一个断点回归设计研究》, 《北京大学教育评论》第 4 期。
- Abdulkadiroglu, A., J. D. Angrist, Y. Narita, P. A. Pathak and R. A. Zaretsky, 2017, “Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago’s Exam Schools”, *American Economic Review: Papers & Proceedings*, 107(5): 240–245.
- Abdulkadiroglu, A., P. A. Pathak and C. R. Walters, 2014, “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools”, *Econometrica*, 80(1): 137–196.
- Betts, J. R., 2011, “The Economics of Tracking in Education”, *Handbook of the Economics of Education*, (3): 341–381.
- Bui, S. A., S. G. Craig and S. A. Imberman, 2014, “Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Achievement”, *American Economic Journal, Econ. Pol.* 6(3): 30–62.
- Calonico, S., M. D. Cattaneo and R. Titiunik, 2014, “Robust Data-Driven Inference in the Regression-Discontinuity Design”, *Stata Journal*, 14(4): 909–946.
- Calonico, S., M. D. Cattaneo and R. Titiunik, 2014, “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs”, *Econometrica*, 82(6): 2295–2326.
- Cullen, J., B. Jacob and S. Levitt, 2005, “The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools”, *Journal of Public Economics*, 59(5): 729–760.
- Davis, A. J. 1966, “The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men”, *American Journal of Sociology*, 72(1): 17–31.
- Dee, T. and X. Lan, 2015, “The Achievement and Course-taking Effects of Magnet Schools: Regression-discontinuity Evidence from urban China”, *Economics of Education Review*, 47: 128–142.
- Ding, W. and S. F. Lehrer, 2007, “Do Peers Affect Student Achievement in China’s Secondary Schools?”, *Review of Economics & Statistics*, 89(2): 300–312.
- Duflo, E., P. Dupas and M. Kremer, 2011, “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”, *American*

- Economic Review*, 101(5): 1739—1774.
- Gelman, A. and G. W. Imbens, 2019, “Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs”. *Journal of Business & Economic Statistics*, 37(3): 447—456.
- Hahn, J., P. Todd and W. Van der Klaauw, 2001, “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design”, *Econometrica*, 69 (1): 201—209.
- Hastings, J. S. and J. M. Weinstein, 2007, “No Child Left Behind: Estimating the Impact on Choices and Student Outcomes”, *NBER Working Paper*, No. 13009.
- Imbens, G. W. and K. Kalyanaraman, 2012, “Optimal Bandwidth Choice for the Regression Discontinuity Estimator”, *Review of Economic Studies*, 79(3), 933—959.
- Lee, D. S. and T. Lemieux, 2010, “Regression Discontinuity Designs in Economics”, *Journal of Economic Literature*, 48: 281—355.
- Liu, M. and J. Jackson, 2011, “Reticence and Anxiety in Oral English Lessons: A Case Study in Mainland China”. In Jin, L. and M. Cortazzi (eds), *Researching Chinese Learners: Skills, Perceptions and Intercultural Adaptations*, Basingstoke: Palgrave Macmillan.
- Marsh, H. W., C. K. Kong and K. T. Hau, 2000, “Longitudinal Multilevel Models of the Big-fish-little-pond Effect on Academic Self-concept: Counterbalancing Contrast and Reflected-glory Effects in Hong Kong Schools”, *Journal of Personality and Social Psychology*, 78: 337—349.
- McCrary, J., 2008, “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test”, *Journal of Econometrics*, 142(2): 698—714.
- Shi, Y., 2020, “Who Benefits from Selective Education? Evidence from Elite Boarding School Admissions”, *Economics of Education Review*, 74: 101907.
- Vardardottir, A., 2013, “Peer Effects and Academic Achievement: A Regression Discontinuity Approach”, *Economics of Education Review*, 36(4): 108—121.
- Vanderhart, P. G., 2006, “Why do Some Schools Group by Ability?”, *American Journal of Economics and Sociology*, 65(2): 435—462.
- Wu, J., X. D. Wei, H. L. Zhang and X. Zhou, 2019, “Elite Schools, Magnet Classes, and Academic Performances: Regression-discontinuity Evidence from China”, *China Economic Review*, 55: 143—167.
- Zhang, H., 2016, “Identification of Treatment Effects under Imperfect Matching with an Application to Chinese Elite Schools”, *Journal of Public Economics*, 142: 56—82.

Can the Elite Classes Improve the Academic Achievement of High School Students? Based on Regression Discontinuity Design of Senior High Schools in a Western City

YANG Jing^{1,3}, YE Xiao-mei², DU Yu-hong¹

(1. Faculty of Education, Beijing Normal University; 2. Graduate School of Education, Peking University; 3. Tianquan Senior High in Sichuan Province)

Abstract: In practice, grouping elite classes is a relatively common resource allocation strategy in high school. However, due to the lack of empirical evidence, its fairness and efficiency have long been controversial. This paper uses the longitudinal data of the students of the 2017 and 2018 cohorts of five high schools in a city in western China, and uses the method of regression discontinuity to test the effectiveness of grouping elite classes from the perspective of improving students' academic performance. The results show that: although the students' academic performance in the elite classes were higher than those in the normal class in the first grade and third grade, the performance gap between the elite class and the normal class did not significantly widen during this period. Further, using the exogenous variable of the cutoff-line of the elite classes in the classes grouping of grade 1, it is estimated that the students' performance difference between entering the normal classes and the elite classes in grade 3, the students whose ability is similar near the cutoff-line. Both the parameter estimation and non-parametric estimation results show that entering the elite classes did not significantly improve the student's academic performance, and no heterogeneous difference between genders and between urban and rural students. From this, it can be seen that the grouping elite classes in high schools are not an effective strategy for optimizing the allocation of educational resources and improving students' academic performance, but are more likely to cure the performance gap between gifted and normal students. In the future, grouping classes in high school need to take into account fairness and efficiency, and promoting the value-added development of different types of students.

Key words: elite classes; academic performance; regression discontinuity design

(责任编辑: 郑磊 责任校对: 郑磊 胡咏梅)