微积分梯度：西班牙裔服务型大学的儿童贫困、数学准备程度与大学一年级表现

JuanB. Guti´érrez 数学系 德
克萨斯大学圣安东尼奥分校

arXiv:2611.29032v1 [stat.AP] 21 Nov 2026

摘要

圣安东尼奥都会区是美国大型都会区中贫困率最高的地区之一。在城市所在的贝克萨尔县，各学区的儿童贫困状况差异很大。我们将该市公立研究型大学——一所广泛招生的西班牙裔服务型院校——的入学者与其生源学校和学区进行关联，并考察他们的准备程度分数和第一年结果。在全美各学区，广泛修读的核心课程（代数 II、几何、生物学）的入学人数与儿童贫困的关联不大；而儿童贫困程度较高的地区，无论在全国范围内还是贝克萨尔县内，微积分、大学先修课程数学和科学课程的入学人数都低得多。在能够关联到生源信息的入学者中，数学准备程度因学校和学区而异。数学准备程度较高的入学者，几乎在所有学科都取得了更高的第一学年成绩。第一学年后未返校的大多数入学者都曾有课程未能顺利完成，而数学是导致此类结果的最大单一来源。在更换专业的学生中，数学是成绩远低于其个人平均水平的学科之一。在同一课程和学期内，对于数学准备程度较低的入学者，较大的数学课程班规模与更多不成功结果相关，这为测试小班教学提出了一项假设。这些结果记录于2014年秋季至2019年秋季，可作为疫情、可选提交考试成绩的招生政策以及生成式人工智能之后变化的参照点。本分析可复现；若要在其他地方采用，则需要制定本地定义并进行关联验证。所有结果都是某一所院校学生群体中的关联；若要确定因果关系，则需要开展设计严谨的研究。

关键词：基础数学；教育公平；儿童贫困；记录链接；西班牙裔服务型院校；可复现研究

数学主题分类：62P25（主要）；62H30；62J05；97B40

1 引言

按2025年7月的人口估算，得克萨斯州圣安东尼奥是美国第七大城市 [1], ；其大都会区的贫困率在2022年为14.2%，在……中位居最高之列
在美国最大的二十五个大都市区中排名第二，按点估计值计算，并且在统计上与休斯敦（14.3%）和底特律（13.8%）没有显著差异 [2]。这座城市在美国学校财政史上也占有特殊地位。得克萨斯州主要通过地方财产税为公立学校提供资金；对这一安排提出的宪法挑战，即圣安东尼奥独立学区诉罗德里格斯案，起源于该市的埃奇伍德学区。最高法院裁定，教育并非联邦宪法规定的一项基本权利，并依据合理性审查标准维持了这一财政制度 [3]。五十年后，贝克萨尔县各学区的儿童贫困程度仍然差异很大，从阿拉莫海茨（7%）到埃奇伍德（39%）（第 3.1 节）。更高的教育

程度与更高的终身收入和更低的失业率密切相关 [4, 5]。学生日后的收入和衡量得出的收入流动性，也因他们就读的学院而有所不同；这些差异既反映了各院校招收的学生，也反映了院校提供的教育[6]。因此，这一地理格局引发的公平问题十分具体：哪些学校和学区将学生送入大学，它们与学生入学时的数学准备程度有何关联；而这种准备程度又与学生的第一年表现有何关联。

我们在得克萨斯大学圣安东尼奥分校（UT San Antonio）探讨这一问题。该校是圣安东尼奥市的公立研究型大学。以下三项事实描述了该校当前的情况；这些情况发生在本文分析的记录所涉及的时期之后。UT San Antonio 在2025年分类中获得了卡内基R1研究型大学认定 [7], ，并于9月与 UT Health San Antonio 合并 2025 [8]。该校入学机会相对广泛：2024年秋季，该校录取了87%的首次申请本科生 2024 [9]。该校是联邦认定的西班牙裔服务型院校，2024年秋季入学学生中，60%为西班牙裔，43%为第一代大学生 [10]。该校广泛的入学机会以及西班牙裔服务型院校的认定，使其成为考察记录在案的学校和学区背景如何关联准备程度与第一年表现的合适研究场所。

我们提出三个问题。第一，在入读该校的学生中，家庭相关指标、生源高中课程选修情况和学区儿童贫困与入学时的数学准备程度有何关联？第二，数学和语言准备程度与学生第一学年所修课程的成绩有何关联？第三，在该校内部，第一学年课程结果与持续就读及专业变更有何关联，班级规模又与成绩有何关联？全国范围内学区层面的课程选修模式为第一个问题提供背景。

支撑这些问题的关系已有大量先例。全国性证据已表明，高贫困学校开设的高阶课程较少，而基础课程的开设则更为普遍 [11]。来自佛罗里达州和得克萨斯州的学生记录以及民权数据收集（CRDC）显示，修读高阶课程的机会在贫困程度和种族方面也存在类似的不均衡 [12–14], ；修读难度较高的课程还与高中毕业及进入四年制大学相关 [15]。加利福尼亚大学圣迭戈分校和得克萨斯州公立大学的关联研究，将家庭背景及生源高中特征与大学学业成就联系起来 [16–18]。其中一项研究涵盖了本校。Fletcher 和 Tienda [17] 利用四所得克萨斯州公立大学（包括德克萨斯大学圣安东尼奥分校）20 世纪 90 年代及 21 世纪初入学者的记录发现，对比就读于同一所高中的学生后，大学成绩中的黑人与白人差距以及西班牙裔与白人差距大多消失。相关证据还来自全国新生调查以及某大型学区的毕业生 [19, 20]。一项涵盖十五所公立大学的全州研究，考察了高中成绩、考试成绩和课程难度（既按绝对水平衡量，也相对于生源学校衡量）与大学学业结果之间的关系 [21]。这些研究证实，大学入学者所处的学校背景很重要。我们的贡献在于测量这一背景中的特定关系。我们的入学者比 Fletcher 和 Tienda [17],研究的入学者晚十多年入学；我们通过可量化的指标——学区儿童贫困程度以及学校记录的课程选修情况——描述生源背景，而该研究则使用各所高中的指标。

课程层面分析也有先例。早期研究比较了数学和语言成绩对各类大学课程的预测贡献，并记录了课程选择和评分方面的差异 [22]。关于招生考试效度的研究进一步表明，为什么必须结合选拔和背景特征来解读预测关联 [23]。我们通过在统一设定下考察 25 个学科的首次修读情况，比较修读同一课程且处于同一学期的学生，并评估在整所生源学校均被排除在样本之外时，学区贫困程度和已测量的学校课程选修情况所带来的额外预测贡献，从而拓展了这项研究。现有记录缺少高中成绩和个人高中成绩单，因此，这项比较并未

确立现有最佳招生或分班模型。

第三个问题同样有先例。学生在了解自己的学业表现时会离开大学 [24]。低成绩预示着专业变更 [25], ，学生在了解自己在科学学科中的表现时会转离科学领域 [26]。大学研究报告称，班级规模越大，成绩越低 [27]。在数学领域，这种关联已被发现对能力较低的学生更强 [28], ，并且已有研究报告称，班级规模对西班牙裔服务型院校中的西班牙裔学生具有负面影响 [29]。我们在一所院校中按学科描述这些关联。我们报告：入学后至少一年仍被观察到在读的、第一学年课程不成功的入学者所占的频率；每门入门学科对未能完成学业的入学者之不成功结果的贡献程度；以及第一学年成绩与记录在案的专业变更之间的关联，在多大程度上取决于哪些入学者仍保持在读。在观察窗口结束前暂时停学后又返校的学生，也计入之后仍被观察到的学生之中。

早期研究曾一次或两次考察这些关联：高贫困学校的课程选修、学校背景与大学学业成就、入学成绩在大学课程中的适用范围、成绩与持续就读、成绩与专业变更，以及班级规模。据我们所知，这是首项将所有这些因素结合起来，针对一所大型高校的学生进行深入考察的研究。院校分析使用同一份数据提取文件，其中保存了2014年秋季至2019年秋季期间在读的68,558名学生的记录。每项分析都使用这些学生中单独界定的样本：准备度和成绩模型使用关联入学者，学业进展、持续就读和专业变更使用固定的参考队列，其他院校描述则使用观察窗口内活跃的所有学生（表 C1）。在同一所院校内开展研究，正是这种深入程度成为可能的原因：记录保存了每名学生的每次课程修读记录，论文中的每个数字都由其程序重新生成。同一选择也限定了结论的范围。院校统计数据反映的是申请该院校、获该院校录取并选择就读该院校的学生之间的关联；关于得克萨斯州或全国的表述则需要其他数据。附录B.5将该院校置于全国完成学业格局中，以便其他院校根据这些发现确定自身的相对位置。

本研究的首要动因是地理因素。学生从课程选修情况和儿童贫困程度各不相同的学校和学区进入大学，因此，对入学准备程度的研究必须衡量输送这些学生的学校和学区。在2020–21年民权数据收集（CRDC）中，约31,100所提供七年级或八年级课程的公立初中里，有39%在该收集周期内未开设代数 I [30]。在得克萨斯州，2013年众议院第5号法案以基础高中课程计划取代了默认课程体系；根据该计划，代数 II 是获得 STEM认证以及达到杰出成就等级的必修课程，而在其他情况下则为选修课程[31, 32]。全国成绩单证据表明，高中阶段达到的最高数学水平是衡量学生迈向学士学位进程的关键指标 [33, 第 30–33] 页。大学本身的地理可及性也因地区而异 [34]。

本研究的第二个动因与时间有关：后续比较需要一个在2020—2025年这一时期之前测量的参照状态。2020年至2025年间，该系统的多个层面同时发生了变化：COVID-19导致教学中断和紧急评分政策，大量招生开始实行考试可选制，得克萨斯州依据众议院第 8 [35], 号法案重组社区学院经费，生成式人工智能也逐渐融入日常课程作业；截至 2024 [36] ，86%的受访高等教育学生表示在学习中使用了生成式人工智能，而截至 2025 [37]年12月，62%的美国初中及以上学生已将其用于家庭作业。这些变化在时间和涉及的总体上相互重叠，因此，仅凭院校记录无法将结果的后续变化归因于其中任何一项。此类记录能够支持的是，将结果与一个在这些变化开始前便已结束的参照状态进行比较。我们在此构建这一参照状态：参考队列中的每一项测量——即入学时间介于2014年秋季和

2015年秋季的首次入学新生——均在2020年秋季之前结束（第2.2节）。

本文结构如下。第2节介绍机构数据库，界定参考队列及其观察窗口，建立地理关联，并界定关联入学者模型和班级规模模型。第3节首先列出论文的五项发现，随后报告学区课程梯度（第3.1节和第3.2节）、关联准备程度模型和第一年表现模型的摘要（第3.3节），以及院校在持续就读与专业变更、相对学科评分和班级规模方面的结果（第3.4节至第3.6节）。第4节综合讨论这些发现及其启示与局限性，并提出一项2020年后比较方案。正文辅以四个附录：定量方法，开篇介绍严谨性与可重复性的记录（附录A）；补充结果，包括完整的关联模型和班级规模模型（附录B）；数据清单、关联验证与治理（附录 C）；以及图3中各专业的角度索引（附录 D）。

2 数据与方法

2.1 机构数据库

机构记录是从大学学生信息系统中摘取的课程教学班粒度数据：每行对应一名学生、一个学期和一个课程教学班（共 1,984,991 行，涉及 92,769 名学生），学期范围为 2009年秋季至 2019年秋季。本文的机构分析使用了在 2014年秋季至 2019年秋季期间至少有一条课程记录的 68,558 名学生（这些学期共 1,236,347 行）；我们将他们称为时间段内活跃的学生。另有一张学生级表，记录其中 61,954 名学生的大学前数学学分。关联字段的覆盖率很高：98.6% 的时间段内活跃的学生记录了生源高中名称，99.96% 记录了家庭邮政编码（附录C.1）。每个与该记录合并的外部数据集都存入一个按数据溯源分层的 PostgreSQL 数据库中，其原始层按原样保存源文件及其 SHA-256 哈希值；来源登记簿则记录每个导入文件的数据集、版本、路径、哈希值、字节数、获取网址和行数。本文获取的档案还包含完整的清单条目，记录获取、许可、模式和谱系的详细信息（附录C.1）。数据溯源依据哈希值和登记簿条目确立，而非文件名或行数。本研究获得了研究院校机构审查委员会的批准，且机构数据摘录本身受到限制（见“伦理审批”和“数据可用性”声明）。附录C收录了完整清单（表 C2）、来源缺陷目录和推断惯例。

2.2 参考队列与观察窗口

这些分析采用不同的人群，因为它们回答的问题各不相同；附录 C 中的表 C1 列出了每个人群及其观察窗口和分析单位。参考队列 包括 $N = 10{,}027$ 名在 2014年秋季至 2015年秋季期间入学的首次入学新生：2014年秋季有 4,935 人，2015年秋季有 4,923 人，2015 年春季或夏季有 169 人。首次入学新生是指其首个记录学期带有注册主任所设首次入学新生类型代码的学生。我们从每名学生的入学学期开始，观察至四年后的同一学期（含该学期）。因此，2014年秋季入学者的观察期截至 2018年秋季，2015年秋季入学者截至 2019年秋季，2015 年春季或夏季入学者则截至 2019 年春季或夏季学期，因此所有队列测量都在 2020年秋季之前结束。对于秋季入学者，观察窗口涵盖前四个学年及其后的秋季学期。所有报告的队列数量均采用这一含边界的定义，第 4.4 节的比较方案也要求 2020 年之后的重跑分析遵循该定义。严格四年版本作为敏感性检验报告。

进阶结果是在观察窗口内注册修读编号严格高于 4000 的课程记录，即高年级课程。入门课程包括数学（教务处课程代码 MAT）、写作（WRC）以及必修一年级研讨课“学术探究与学术研究”（AIS）。在队列表中，接触某一学科是指在观察窗口内任何时候修读该代码下的任意课程；仅在第 2.4 节的关联入学者模型中，才限定为第一学年。该队列包括 1,190 名被标记为协调招生计划（CAP）的学生，他们计划在第一学年后转学：其中 24.5% 在该院校修读了高年级课程，而其余 8,837 人中这一比例为 54.0%。队列数量均针对完整队列报告；未带该标记的学生中，数学优势比为 1.42。

观察窗口敏感性与旧版终点规则。 观察窗口敏感性已针对源系统以只读方式重新计算。限制为严格四学年后，达到高年级的比例从 50.5% 降至 47.4%（10,027 人中有 4,755 人），而入门课程的优势比从 1.46 变为 1.60（MAT）、从 1.39 变为 1.45（WRC），以及从 0.23 变为 0.26（AIS）。现报告边界情况，以便明确估计目标的实际实施方式。少于 10 名队列学生修读课程编号恰为 4000 的课程，并因严格不等式而被排除；67 名学生修读研究生编号课程（5000 及以上），并被计入；另有 34 名学生的课程记录日期早于其入学学期，而查询因未设置下限而将其纳入。专业变更标记比较观察窗口内首次和最后一次记录的专业：27 名队列学生缺少这两项记录中的一项，而公布的 43.0% 将其计为发生变更（仅计入已确认的变更时，完整队列中的比例为 42.7%，在 10,000 个完整案例中为 42.8%）。底层查询未在同一学期内设置平局裁决规则；该队列中没有学生在观察窗口的第一学期或最后学期同时拥有两个专业，因此该标记不取决于记录顺序。这些是最初建立队列表时采用的规则，为保持公布数量的可比性而予以保留；队列标记应按这些规则所定义的含义解读。附录 B.7 根据入学界定规则重复了一项比较，其中仅计入已确认的变更。这些计数相对于 $N = 10{,}027$ 均很小；观察窗口边界和优势比计数已作为机器可读构件（cohort window audit.json）与表格一并重新生成。

机构描述性统计可能采用完整观察窗口（2014年秋季至2019年秋季所有在读者，包括入学者和转学生）；每个图都会说明其描述的总体。准备程度的关联模型采用2014年秋季至2019年秋季的秋季首次入学者。一年级成绩模型和班级规模模型采用2014年秋季至2018年秋季的入学者，这些人的第一学年记录完整（第 2.4 节）。所有院校总体均按范围说明纳入研究范围，详见第1节。

2.3 地理关联链

情境测量通过一条经过审计、分为两个分支的链条传递给学生。在学区分支中，我们首先将每名学生记录的高中名称与美国国家教育统计中心（NCES）的学校标识符进行匹配。候选学校在按城市和邮政编码划定的地理区块内进行比较；匹配评分方法见附录A。分配的学校标识符前七位用于识别学区（即其地方教育机构标识符 LEAID）。我们使用学区标识符关联美国人口普查局的小区域收入与贫困估计（SAIPE）中的儿童贫困数据，并使用学校标识符关联民权数据收集（CRDC）中的课程注册指标。这些基于标识符的连接是精确的。第一步中学校是否分配正确，则是另一个验证问题。社区分支将家庭邮政编码解析到邮政编码制表区（ZCTA）层级的机会地图集背景数据 [38], ，这些数据由人口普查分区按人口权重汇总而成；所得变量属于生态层面的数据，并在模式中标记为此类变量

。名称解析对照表覆盖了在得克萨斯州高中名称一栏有记录的 60,045 名在校学生中的 90.7%。学生级表包含 61,954 名在校学生，其最终处置情况为：49,099 人匹配成功，6,945 人就读于得克萨斯州以外的高中，2,749 人无法解析，2,058 人就读于 NCES学校总体之外的私立学校，859 人未记录学校，244 人在审核中被剔除（表C3；学生级表缺少 6,604 名在校学生，见附录C.1）。在样本中，无法解析的学生比匹配成功的学生更常拥有双重注册学分和佩尔助学金。以生源学校为聚类单位时，这些差异处于学校间差异的范围内，因此审计无法确定链接是否在这些标记上存在选择性。无法解析学生的学区贫困数据不可观测，因而其对学区关联结果造成的任何偏差方向都未知；选择审计见附录C.5。分配的标识符与全国学校总体相符，表明这些是真实存在的学校；但在名称相似的候选学校中选中正确学校的比率，尚未通过经裁定的样本进行测量。每次使用这条链时，均受两项范围规则约束：CRDC 快照是特定周期的学校报告，绝不视为学生接触史；学校点位绝不视为服务区。

2.4 关联入学者模型

关联模型针对同一总体提出两个问题：哪些入学特征与入学时的数学准备程度相关，以及哪些特征与各学科的第一年表现相关。分析计划（变量、结果判定规则、模型、假设，以及判定假设获得支持的阈值）在估计任何模型之前就已撰写并冻结，且分析流程首先在打乱后的结果上端到端运行。预先设定为内部文件：项目保留该方案以及方案和代码的 SHA-256 记录，但没有外部注册机构为其加注时间戳。冻结后所作的更改均记录在仅追加列表的偏离项中，并在报告相关结果时逐一说明。方案与代码一并提供。本小节说明总体、测量指标和比较；附录A.5给出完整设定。

总体。 所有在2009年秋季至2019年秋季首次秋季入学者（28,209名学生），即机构描述性统计的观察窗口。关联样本 包括其中生源学校能够在交叉映射表中确定、且学校和学区背景信息完整的23,489名学生；其中23,478名同时具有入学成绩。表现模型使用截至2018年秋季的队列（19,874名关联入学者），这些学生的第一年观察期完整。CAP入学者予以保留，且该标记变量进入每个模型。该方案最初针对2009年秋季至2019年秋季的入学者撰写，并首先在这11个队列上运行。在那次运行之后，研究采用了从2014年秋季开始的限制，其结果当时已经知晓，原因是较早入学届的院校条件有所不同。该限制作为方案中的偏离项（D1）予以记录，方案中的其他内容均未因此改变。本文仅报告2014年秋季至2019年秋季期间在读学生的情况，因此不提供较早入学届的结果。因此，受限分析是在方案冻结后作出的总体选择，报告该分析的表格和图也对此予以说明。

准备情况与表现。 入学准备程度以入学数学成绩衡量：院校记录的 SAT 和 ACT 数学分科成绩最高值，按现行 SAT 量表计分。在每个入学队列内，我们先减去队列均值，再除以队列的标准差，因此一个单位相当于一个标准差的成绩（合并各队列后为 71 点）。语言成绩也以相同方式标准化。第二项指标为标准化数学成绩减去标准化语言成绩，用来描述入学者在两项测试中的相对位置。第一年表现根据低年级课程的首次修读情况衡量，涵盖入学秋季学期及随后的春季和夏季

课程，采用两种形式：按 4 分制计算的成绩，以及 DFW 指标（D 等、F 等、退课或无学分结果）。记录中没有高中平均成绩、班级排名或成绩单。

预测变量和模型。 预测变量描述了四个方面：家庭资源（学生是否申请助学金、申请者的收入排名以及第一代大学生身份）、人口统计特征（性别和族裔）、生源学区的儿童贫困率，以及生源学校记录的课程注册情况（每名学生修读微积分、大学先修数学、AP科学和双重注册课程的情况）。学校的 高阶数学指数 是其标准化微积分和大学先修数学比率的平均值。模型分为三类，均使用这些预测变量。对学区和学校设置随机截距的模型，展示了准备程度的差异有多少存在于学区之间、学校之间，以及同一学校入学者之间。对于至少有1,000名关联入学者进行首次修读的25个学科中的每一个，采用包含课程与学期固定效应的回归模型，将成绩与两项入学成绩及其他特征联系起来，从而把学生与同一学期修读同一课程的其他学生进行比较。最后，对未纳入拟合的生源学校进行预测，以衡量每组预测变量对准确度的提升程度。

预先确定了两个假设。H1：在控制家庭资源后，学校的高阶数学指数与数学与语言成绩差呈正关联。H2：在控制语言成绩后，数学成绩与定量学科以外的第一学年成绩呈正关联。在估计之前，有八门学科被归为定量学科，其中七门的入学者数量足以建立模型：数学、化学、经济学、统计学、物理学、计算机科学和会计。每项估计都是以进入该院校为条件的关联。

班级规模模型。 方案的预先规定补编在对真实成绩拟合任何课程内模型之前已冻结，该补编考察同一课程中班级规模较大的入学者是否会取得较低成绩。各行数据为2014年秋季至2018年秋季所有秋季首次入学者在超过10名学生的面对面授课班级中的首次修读记录（177,289次首次修读，23,766名入学者）。班级规模统计该教学班中登记的每名学生，无论其属于哪个入学队列。模型使用班级规模的以 2 为底的对数，因此，系数表示班级规模加倍所对应的差异。

四个嵌套模型依次逐步收紧比较范围：该学科的所有教学班；课程与学期固定效应；表现模型中的学生协变量；以及教师固定效应，后者构成主要模型。因此，主要模型比较的是在同一学期修读同一门课程的入学者，同时控制其已测量特征，并剔除授课教师之间的平均差异。模型仍保留某位授课教师在某门课程或某个学期中特有的任何差异，并将课程班分配视为其本来的行政流程。第五个模型加入班级规模与入学数学成绩的交互项；另一个加入学生固定效应的汇总模型，则将每名学生与其在其他课程中的表现进行比较。25个学科中有20个拥有足够的学生数量和班级规模差异，因而可纳入分析；纳入规则见附录 A.6。数学方面预先设定了两个假设。H3：在主要模型中，班级规模越大，成绩越低。H4：对于入学数学成绩较低的入学者，这种关联更强。

对打乱的成绩进行试运行后发现，在这种设计中，采用按授课教师聚类的标准误进行常规检验，拒绝原假设的频率过高。因此，班级规模项的区间和 p值来自针对授课教师进行的野生聚类自助法 [39]；包含 758 位授课教师的同一学生模型则保留聚类区间（附录 A.6）。随后，一个情景将每个学生人数超过 45 人的教学班拆分为

人数相等且每班不超过 45 人的教学班，并采用各学科的主要系数。45 是该情景设定的阈值：记录中没有招生上限字段，而 45 是微积分 I 和 II 讲座班级规模观测值的第90百分位数（中位数为 36，没有任何班级超过 50 人）。只有在以下三个观察性模型无法证实的条件下，该情景才描述一种干预：班级规模会导致所估计的差异；在拆分范围内，关联与规模的对数呈线性关系；新增教学班的授课方式与现有教学班相同。

2.5 定量方法

读取图1和图2需要一个估计量。学区课程梯度 从学区内某门课程的入学率开始计算：选课学生人数除以该课程所在年级的估计入学人数。对于高中课程，这些年级为9至12年级；对于在7至8年级修读的代数 I，则为这两个年级。分母反映相关年级学校总体的规模，但无法衡量哪些学生已满足该课程的先修条件。梯度是该入学率对学区儿童贫困率进行回归所得的最小二乘斜率，并以相同的入学人数加权。由于各课程的基准入学率差异很大，每个斜率都以相对尺度报告：儿童贫困率每增加十个百分点所对应的入学率差异，占该课程全国平均值的百分比。区间为95%区间，通过对州进行自助法重抽样得出（5,000个重抽样样本）。

图4中的年级偏差在其使用处定义（第 3.5 节），班级规模系数则与其模型一同定义（第 2.4 节）。附录A完整说明了所有估计量，包括记录链接得分、专业变更图的布局、班级规模曲线的平滑处理，以及关联入学者模型和班级规模模型的设定与推断。代码可用性声明中列出了实现脚本的名称。

3 结果

结果支持五项发现。以下分别列出每项发现及报告该发现的小节。

1. **课程选修与学区贫困程度** （第 3.1 节和第 3.2 节）。儿童贫困程度较高的学区，其微积分以及 AP 数学与科学的入学率要低得多。相比之下，代数 II、几何和生物学的入学率随学区贫困程度的差异很小。这些模式在全国范围内 [11] 以及贝克萨尔县各学区均有体现。入学率根据相关年级学生的估计人数计算。它们衡量的是有多少学生选修这些课程，而不是学校是否开设这些课程。

2. **课程体系中的准备程度与成绩**（第3.3节）。在控制语言成绩和其他纳入分析的特征后，在同一学期修读同一课程的学生中，数学成绩越高，其成绩越高。该系数在定量学科中更大。在定量学科之外，数据无法证明数学成绩的关联大于语言成绩。

3. **背景与预测**（第3.3节）。入学者的数学成绩因其生源学校所在学区的贫困程度和高等数学入学情况而异。然而，纳入入学成绩后，再加入这些背景变量对提升大学成绩预测的作用微乎其微。即使在对来自未参与模型训练的学校的学生进行预测测试时，结论依然如此。这一发现并不意味着学校背景无关紧要，也无法说明准备程度差异是由什么造成的。

4. **第一年未成功结果与持续就读**（第 3.4 节）。入学一年或更久后没有注册记录的大多数入学者，在第一学年曾取得 D、F、退课或无学分结果。与其他任何学科相比，数学课程造成了更多此类结果。在所考察的每一门入门学科中，结果不成功的入学者之后被观察到仍在注册就读的可能性较低。这种关联主要集中在多门学科都出现不成功结果的入学者身上。这些比较并不能证明课程结果导致学生离校。

5. **班级规模**（第3.6节）。早期大学研究考察了因学生而异的班级规模关系，以及同一课程和授课教师内部的比较 [40, 41]。在这里，在同一课程和学期内，班级规模与数学成绩之间较小的平均关联，与班级规模和准备程度之间的交互项并存：入学数学成绩较低时，较大的教学班对应更多 DFW结果；De Paola 等人 [28] 报告了意大利大学数学课程中的相同方向。此发现为该院校开展设计评估提供了依据，但不能证明缩小班级规模会产生效果。

第 3.4 节和第 3.5 节还介绍了修读数学课程的学生中发生的专业变更。经审计的记录链接（第 2.3 节；附录 C ）和 2020 年前参考量（附录 B.10 ）为这些发现及后续重复验证提供了支持：根据明确说明的定义，这些资料可以用于比较不同人群和结果，并同时报告链接失败和选择限制。

学区分析使用已发布汇总数据，分析单位为学区。根据第 2.4 节的规则，学生记录在两个地方进入模型：关联入学者模型（附录 B.4）和班级规模模型（第3.6节）。所有结果都只是关联，本节中的任何内容都不支持因果解读。

3.1 学区微积分选课人数、儿童贫困与收入

图1将贝克萨尔县各学区置于全国微积分选课人数横截面中，并分别与学区儿童贫困（A面板）和学区家庭收入（B面板）进行比较。微积分是高中加速数学课程序列的最后一门课程；第 3.2 节发现，在所比较的课程类别中，微积分选课人数与学区儿童贫困的负向关联程度位居前列。选课人数按每名高中生计算：以民权数据收集（CRDC）中的微积分选课学生人数除以该学区各校估算的高中在校生人数。

儿童贫困率较高的学区，微积分选课人数较少。在全国学区总体中，每100名高中生中有4.6人选修微积分。儿童贫困率高出10个百分点的学区，平均每100名高中生中少1.7名微积分选课学生，相当于全国平均入学率的36.9%（按入学人数加权的斜率 -0.171 每单位贫困率；按州聚类的标准误为0.020）。这是对学区的横截面比较，并未估算某个学区的贫困率发生变化时，其选课人数会如何变化。未加权拟合结果与此一致（-0.179，标准误为0.019）；在得克萨斯州，斜率较小（-0.083，异方差稳健标准误为0.023；贫困率每增加10个百分点，相当于得克萨斯州平均值的24.7%）。贝克萨尔县各学区的儿童贫困率差异很大，从阿拉莫海茨独立学区（ISD；儿童贫困率为7%）到埃奇伍德独立学区（39%）不等；这两个学区分别代表了第1节所述学校财政史的两极，也位于全国分布的两端：阿拉莫海茨每100名高中生中有12.0人选修微积分，是全国数值的2.6倍；埃奇伍德则为0.7人，是全国数值的0.14倍。该县最大的两个学区——北区独立学区和 North East ISD——的数值略低于全国

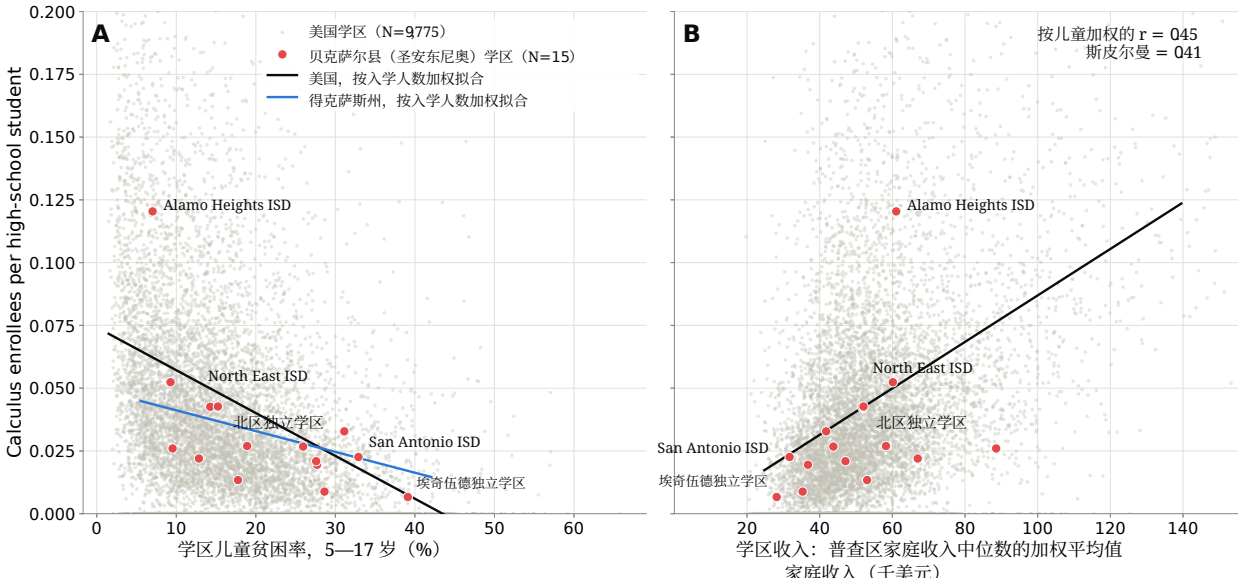


图1：共同样本框中9,790个学区的每名高中生微积分选课人数与（A）学区儿童贫困率（SAIPE 2017）及（B）学区收入之间的关系；学区收入为普查区家庭收入中位数的加权平均值（美国社区调查 2012–16年，随机会地图集提供的普查区协变量；普查区权重为儿童队列人数乘以土地面积占比）。（纳入条件：匹配成功、至少有200名5–17岁儿童、报告了高中课程序列，且课程项目未被编码为缺失。）分母为估算的高中在校生人数：开设9至12年级课程的学校的总注册人数，按年级范围折算。（A）黑线表示美国按入学人数加权的最小二乘拟合线，斜率为 -0.171每单位贫困率（按州聚类的标准误为0.020）；按儿童贫困率每增加10个百分点计算，该斜率相当于全国平均值（0.046）的 -36.9%；蓝线表示得克萨斯州内部的同类拟合线（860个学区）。（B）有收入数据的9,775个学区；按儿童人数加权的相关系数为0.45（斯皮尔曼相关系数为0.41）。红色表示得克萨斯州贝克萨尔县的学区（样本框内共15个，标注了具有代表性的学区）。2,321个学区报告微积分选课人数为零，另有112个超出图示范围的学区未予显示。分析单位为学区；入学人数统计的是一个周期内的选课席位，不包含课程完成情况或结果内容。

数值（为其0.92倍），而 San Antonio ISD 的数值则是全国值的一半。由于公开使用文件经过披露调整，单个周期中单个学区的数值存在较大波动，仅作示例。

收入面板呈现出相同的排序。其收入指标是各普查区家庭收入中位数的加权平均值。普查区数值取自随机会地图集提供的美国社区调查（美国社区调查）2012–16年协变量 [38]。我们汇总学区内各普查区的数据，并以机会地图集中该普查区的儿童队列人数乘以其土地面积在该学区内所占比例作为权重 [42]。附录A列出了边界文件及其数据版本。我们将普查区家庭收入中位数的这一加权平均值称为学区收入；它不同于学区家庭收入中位数。在贝克萨尔县，收入排序从埃奇伍德和圣安东尼奥学区（约 30,000 美元）到阿拉莫海茨学区（超过 60,000 美元）依次递增（在全国范围内，各学区收入与儿童贫困的无权重相关系数为 -0.72）。每名高中生的微积分选课人数随学区收入增加而上升：按儿童人数加权的相关系数为 0.45（斯皮尔曼相关系数为 0.41）。

在 9,790 个学区中，有 2,321 个学区报告该周期没有微积分选课人数。入学人数指标的限定条件同样适用于两幅图，并在第 3.2 节末尾说明。

3.2 不同课程类别的贫困梯度

图2将比较扩展至民权数据收集（CRDC）学校文件中报告的 15 个课程类别。A面板采用图1中的 9,790 个学区，B面板则将同一指标应用于贝克萨尔县的 15 个学区。每门课程均以其对应年级的估计入学人数为基准；每个斜率采用第 2.5 节所述的相对尺度：儿童贫困率每增加 10 个百分点，入学人数随之变化的幅度，以该课程全国平均值的百分比表示（方程 (A2)）。在这一尺度上，三门选修人数众多的核心课程估计贫困斜率较小，且其区间包含零：生物学（-0.6%，95% 自助法区间 -2.3 至 $+1.7$）、代数 II（-0.7%，-4.7至 $+1.9$）和几何（$+1.1$%，-1.0 至 $+3.2$）。这一结果反映的是学区平均值，并不能证明同一学区内各学校的入学机会或入学人数相同。

负向关联集中在加速课程序列中。大学先修课程数学（36.9%）、微积分（36.9%，区间为28.7至46.2）和大学先修课程科学（35.9%）的降幅最大。按十分位组划分的情况呈现出相同的对比，无需进行线性拟合：在高中学生中，来自最低贫困学区的十分位组（儿童贫困率低于5.6%）的微积分人均选课人数是全国水平的1.93倍；而来自贫困程度最高的十分位组（高于28.6%）的学生，这一数值为0.46倍。相比之下，生物学对应的数值为1.00和0.99。任何课程的大学先修课程选修人数（24.5%）、在7至8年级修读代数 I、高等数学（14.9%）和物理学（13.3%）位于排序中段。对于在7至8年级修读代数 I 的学生，主周期的点估计显示，这些年级的学生人均选课人数下降16.4%，区间包含零（-4.9 至34.0）；在2015–16年周期中，估计降幅为25.9%，区间不包含零（20.6至31.3）。在9至10年级修读代数 I 的估计值略为正向（$+6.5$%，区间 -2.0至 $+13.8$），这符合预期：在高中前完成该课程的学生较少。大学先修课程各类别彼此重叠，也与微积分重叠，因此最上方几行是对高级课程选修情况的相关测量，其排列仅用于展示。这一模式与数学课程序列的结构一致：在9年级前完成代数 I 的学生，到12年级时可以修读微积分。对于这条加速路径，路径末端的选课人数在儿童贫困率较高的学区更低；路径起点的估计方向相同，但精确度较低；而序列中段的选课人数则呈现出近乎为零的关联。

圣安东尼奥所在的贝克萨尔县也呈现出相同的对比。图2的 B 面板重复展示了贝克萨尔县十五个学区的情况。我们按儿童贫困程度排列这些学区，并将其合并为五组，每组至少占该县高中在校生估计人数的 10%：阿拉莫海茨、萨姆·休斯敦堡、伦道夫机场、拉克兰和东北独立学区（儿童贫困率为 7.0% 至 14.3%）；北区独立学区（15.2%）；贾德森和东中央独立学区（17.8% 至 18.9%）；西南独立学区、南区独立学区、南圣安东尼奥和哈兰代尔独立学区（25.9% 至 28.6%）；以及萨默塞特、圣安东尼奥和埃奇伍德独立学区（31.1% 至 39.1%）。贫困程度最低的两组中，修读微积分的比例约等于全国比率（分别为 1.02 倍和 0.92 倍）；其他三组则不到全国比率的一半（分别为 0.38 倍、0.40 倍和 0.45 倍）。各组的代数 II、几何和生物学修读比例均接近全国比率（介于 0.81 倍至 1.26 倍之间）。该县的微积分梯度为：儿童贫困率每增加十个百分点，修读比例便下降相当于全国平均值 32% 的幅度（95% 区间为 11% 至 63%；此区间来自对十五个学区进行重抽样的自助法），而全国相应降幅为 37%。大学先修课程数学（34%）和大学先修课程科学（32%）的点估计也接近全国数值，但区间包含零；三门普遍修读的核心课程的梯度则都在零的七点范围内。

县级面板有两点需要说明。15 个学区的区间较宽；在梯度最陡的三门课程中，只有微积分的区间不包含零。并且，局部下降仅出现在微积分和大学先修课程中：在该县，高等数学呈正梯度（$+16$%，区间为 0.2 至 35），而 7至8年级的代数 I 和双重注册未显示任何

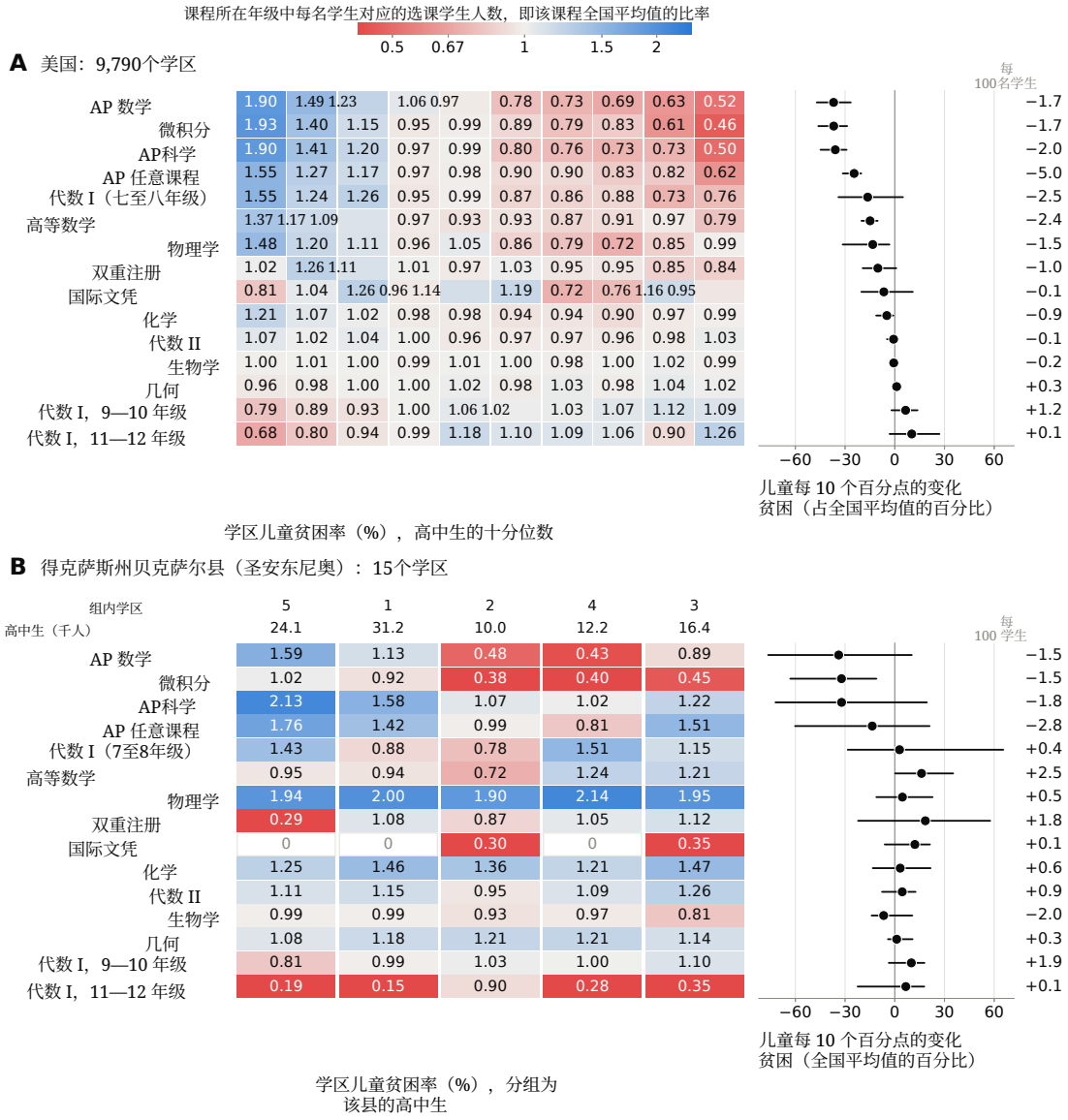


图2：按课程类别划分的课程选修与学区儿童贫困（CRDC 2017–18年；SAIPE 2017年）。每项入学人数统计均除以该课程所在年级的估计入学人数（9至12年级，或7至8年级中代数 I 所在的7、8年级），每个单元格则表示该比率与同一课程全国比率之比（1.00为全国比率；高于该值为蓝色，低于该值为红色；采用对数色阶）。（A）美国，即图1中的9,790个学区。左：按学区儿童贫困程度排序的高中生，每十分之一对应的学生。右：按儿童贫困程度计算的比率加权最小二乘斜率，以每增加十个百分点的儿童贫困所对应的全国平均值百分比表示，95% 区间来自州聚类自助法（5,000个重抽样样本）；右侧数字将各斜率重新表示为每100名学生的选课学生人数。两个面板中的各行均按该梯度排序。（B）得克萨斯州贝克萨尔县15个学区的相同展示。学区按儿童贫困程度排序，并合并为5组，每组至少包含该县十分之一的高中生。斜率根据这15个学区计算，95% 区间来自对学区进行重抽样的自助法。标记为0的单元格表示没有入学人数。学区是分析单位；入学人数统计的是一个周期内的座位数，不包含完成或结果内容。

下降。

附录 B.1 报告的检查结果表明，这一对比依然成立。在所涵盖的三个 CRDC 周期（2015–16年、2017–18年和2021–22）中，三门广泛开设的核心课程的梯度均保持在零上下四点以内，而微积分梯度分别为30.7%、36.9%和45.3%。加入州固定效应、采用范围更广的学区样本，以及仅纳入年级范围无需估算的学校，这些处理下微积分梯度均处于37.0%至37.5%之间，广泛开设的核心课程梯度则均在零上下三点以内。在得克萨斯州，第5号众议院法案规定，除要求修读代数 II 的专业方向外，代数 II 均为选修课；该州每名高中生的代数 II 选课人数不存在贫困梯度（$+0.4$%）。

相对尺度以每门课程自身的平均值为分母计算梯度，因此，基数较小的课程看起来梯度可能很陡，但涉及的学生人数却很少。因此，图2还以选课人数列出各梯度：在某门课程所覆盖的年级中，儿童贫困率每增加10个百分点，每100名学生的选课人数相应变化多少。按选课人数计算，降幅最大的是任意课程的大学先修课程选修人数（每100名高中生减少5.0人；按州聚类的区间为4.1至5.9人）、7至8年级的代数 I（每100名7至8年级学生减少2.5人，区间包含零，-0.8至5.7人）以及高等数学（减少2.4人；区间为1.6至3.1人）；其次是 AP科学（减少2.0人）、微积分（减少1.7人）和 AP 数学（减少1.7人）；广泛开设的核心课程变化不超过0.3人。以学生人数衡量，早期修读代数 I 和高等数学的点估计至少与微积分的估计一样大。

学区收入对课程的排序方式相同。按每名高中生计算，收入与入学人数之间按儿童人数加权的相关系数，在广泛开设的核心课程中介于 -0.06和 0。06 之间；而高等数学为 0.28，微积分为 0.45，AP 数学和 AP科学均为 0.48（9,775 个学区）。收入的地理分布与课程体系的高阶部分相对应，却很难说明广泛开设的核心课程情况。

需要注意四点。入学人数统计反映一个周期内的课程席位，因此无法说明学生在哪个年级修读课程、是否完成课程或取得何种结果。学区层面的汇总数据可能掩盖同一学区内各学校之间的分层情况。高中生这一分母是根据学校入学人数总计和年级范围估算得出的，并已通过上文所述无需估算的学校进行核验。此外，7至8年级的代数 I 数据仅涵盖样本框中开设这些年级的学区；为中学学区输送生源的小学学区不在样本框内。

3.3 入学时的准备情况与一年级成绩

上面两幅图描述的是学区。其余结果描述的是该院校的学生。本小节说明第 2.4 节的关联入学者模型得出了什么结论；附录 B.4 完整报告了这些模型。

入学时的数学准备程度因生源学校和学区而异。 在2014年秋季至2019年秋季的23,489名关联入学者中，入学数学成绩的差异有22%存在于不同生源学校和学区之间，78%则存在于同一学校的入学者之间。来自儿童贫困程度较高学区的入学者入学时成绩较低；学区儿童贫困指数每提高一个标准差，成绩低10.4点。来自高等数学入学人数更多的学校的入学者入学时成绩较高；学校高等数学指数每提高一个标准差，成绩高9.7点。成绩的一个标准差为71点。H1关注数学成绩相对于语言成绩的表现，结果支持这一假设，但系数较小（指数每提高一个标准差，成绩相差$+0.027$ 个标准差；区间为0.008至0.046）。该相对指标的差异几乎全部存在于同一学校的入学者之间（97.5%）。因此，学区贫困程度和学校背景与两项成绩的水平相关，但与两者之间的差异仅存在较弱关联。

入学者具有　数学准备程度更高的学生在几乎所有学科中都取得了更高的一年级成绩　y

学科。在同一学期修读同一门课程且语言成绩相同的学生中，数学成绩在全部 25 个学科中的系数均为正，并且在其中 24 个学科中通过了错误发现控制。分数每提高一个标准差，定量学科以外的绩点就高出 0.14（区间为 0.109 至 0.171），这支持 H2；在定量学科中，绩点则高出 0.26（0.197 至 0.328）。在数学课程中，分数每提高一个标准差，绩点就高出 0.38，DFW概率则低 10.6 个百分点。在定量学科以外，语言成绩的系数与之相近（0.11 个绩点），两种成绩之差的区间包含零。在定量学科中，语言系数较小（0.06）。

纳入入学分数后，学区贫困程度和学校课程选修情况对成绩预测的贡献很小。在留出整个生源学校进行检验时，完整的预测变量集使各学科预测成绩的平方误差中位数降低了 8.6%。入学分数和大学预修学分获取途径贡献了其中的大部分降幅（0.086 中位数降幅中的 0.056）。学区贫困程度使误差中位数变化 $+0.0004$，学校变量使误差变化 -0.0015。这一结果仅涉及使用现有变量进行预测。学区贫困程度在若干学科中仍具有负系数，而这些分析并未估计贫困与成绩之间的关联有多少是由学校课程供给或准备程度导致或中介的。

这些结果描述的是 2014年秋季至 2019年秋季的入学者。作出这一限制，是因为方案首次运行时纳入了 11 个入学队列，且当时已知其结果（第 2.4 节）。

3.4 一年级数学、持续就读与专业变更

其余结果描述了该院校的情况；现有证据表明，学业风险集中在一年级数学上。几乎所有学生都接触到了数学。在参考队列（第2.2节）中，91.0%的入学者在观察窗口内至少选修了一门数学课程，50.5%修读了高年级课程，43.0%有记录显示发生了专业变更（附录B.9）。一份经院校批准后可供查阅的校内机构报告显示，约四分之三的新生其学位计划包含定量要求；数学系每年开设约12,000个课程名额，其中基础服务课程约占95% [43]。以下结果采用两个总体：参考队列，以及整个观察窗口内（2014年秋季至2019年秋季）所有入学者。

全国记录也呈现出同样的集中趋势。在全国四年制学院和大学的数学系中，93%的数学课程选修集中在大学预科、入门级或微积分级别的课程，这一层级还包括线性代数和微分方程 [44, 表 S.2]。全国研究描述了补习数学中的相关模式。公立两年制和四年制院校的新生中，补习课程修读十分普遍 [45], ；而在被转介至发展性数学课程序列的社区学院学生中，只有少数人能完成该序列 [46]。一项随机试验发现，被评估为需要补习代数的学生若被分配到附有每周辅导课的大学水平统计学课程，其通过率高于被分配到补习课程的学生 [47]。

我们首先关注哪些人留了下来。比较对象是参考队列中未标记为参加协调招生计划的 8,837 名入学者；根据该计划的设计，标记为参加该计划的 1,190 名入学者会在第一学年后转学，其中 64.8% 的人没有后续记录。如果入学者的记录显示，在入学学期一年或更久之后、且不晚于观察窗口结束时，仍修读了某门课程，我们就称其为*之后仍在校就读* 。曾暂时停学后又返校的入学者也计为之后仍在校就读。在这 8,837 名入学者中，有 2,006 名（22.7%）未被观察到之后仍在校就读。记录无法说明这类入学者是离开了高等教育，还是转学到了另一所院校，因此他们被视为离开了本院校。本小节的比较

未经调整，是在分析计划冻结后补充的。

第一学年未成功结果很常见，而且与后来未被观察到有关。我们将第一学年入门课程中每一次以D、F、退课或无学分结果告终的首次尝试，计为一次 DFW结果。入学者共有 14,145 次此类结果，占其 67,838 次首次尝试的 20.9%；5,267 名入学者（在有过此类尝试的入学者中占 59.7%）至少有过一次。该结果发生在第一学年，因此早于与之进行比较的后续入学情况。后来未被观察到的入学者比例随 DFW 结果数量增加而上升：没有此类结果的入学者中占 12.7%，有一次的占 14.6%，有两次的占 19.4%，有三次的占 30.9%，有四次或更多的占 52.6%。在后来未被观察到的 2,006 名入学者中，1,553 名（77.4%）至少有过一次 DFW结果。

数学是这些结果的最大单一来源（见附录 B.7 中的图B6）。三个百分比描述了数学所占的位置，每个百分比都有各自的分母。首先，数学课程占所有一年级入门课程 DFW 结果的 23.3%，而占比第二高的学科为 12.8%（若计入一年内的重复修读，则数学占 25.4%）。其次，数学的 10,138 次首次尝试中，有 32.5% 以 DFW 告终，在比较的学科中比例最高。第三，在后来未被观察到的入学者中，数学占 DFW 结果的 20.7%，高于任何其他学科（占比第二高的学科为 13.9%）；在持有 DFW 的入学者中，70.8% 在数学课程中持有 DFW。这些百分比描述的是课程结果。另一个单独的比较描述学生：在至少有一次 DFW 的 3,164 名数学课程入学者中，34.8% 后来未被观察到；而没有 DFW 的数学课程入学者中，这一比例为 14.8%，相差 20.0 个百分点（逐点 95% 区间为 18.0 至 21.9）。

在其他入门学科中没有 DFW 的入学者里，数学方面的差异要小得多：14.2% 对 12.7%，相差 1.6 个点，其区间包含零（−1.3至 4.5）。在同时在其他学科中获得 DFW 的数学入学者中，这两个比例分别为 40.3% 和 18.7%。因此，记录显示，后来未被观察到的入学者普遍存在课程作业未成功完成的情况，同时也无法确定数学是否在他们离校过程中发挥了独立作用。按照其他课程的结果进行划分只是一种描述：这些结果与数学结果的成因可能相同。在十九门拥有足够入学者的入门学科中，每一门都出现了同方向的差异，其中十四门的差异大于数学学科（该图的 B面板和表 B3）：一年级研讨课或写作中的 DFW，与后来未被观察到的情况相关的程度，比数学中的 DFW 更高。这些区间均为逐点区间，且都位于零之上；这些比较属于探索性分析，未根据学科数量进行调整。

离校的原因有很多，而记录中能体现的原因寥寥无几，因此这里陈述的是占比。记录所能支持的结论是：大多数后来未被观察到的入学者都取得了 DFW 结果，而数学造成的这类结果多于任何其他学科。这些比较并不能证明课程结果导致了离校。

专业变更将一年级数学与继续就读学生的学位学习其余阶段联系起来。我们在参考队列中考察了观察窗口内，一年级入门课程首次修读获得 DFW 与有记录的专业变更之间的关系。比较的两个专业分别是记录中的第一个和最后一个，因此无法根据课程结果确定专业变更发生的时间。在有数学课程首次修读记录、且入学至少一年后仍可观察到在读状态的 6,390 名入学者中，2,196 名获得 DFW 的学生中有 65.2% 更换了专业；4,194 名未获得 DFW 的学生中这一比例为 54.6%，相差 10.6 个百分点（95% 区间为 8.1 至 13.1）。在所有有此类首次修读记录的 8,892 名入学者中，无论其之后是否继续入学，这两个比例分别为 45.8% 和 42.9%，相差 2.8 点数（0.7 至 5.0）。因此，这种关联很大程度上取决于入学者在第一学年后是否仍处于可观察范围内：离校的入学者记录专业变更的机会较少。数学是获得 DFW 的入学者人数最多的入门学科。数学和物理学是仅有的两个学科，其差值在两个人群中均为正，且区间不包含零；物理学中较晚仍可观察到的入学者里有 127 人获得 DFW，而数学中有 2,196 人（附录 B.7 中的表 B4）。

两个人群中的差值均为正，且区间不包含零；物理学中较晚仍可观察到的入学者里有 127 人获得 DFW，而数学中有 2,196 人（附录 B.7 中的表 B4）。按学生计算，较晚仍可观察到的入学者中，计算机科学和化学的差值更大（分别为 17.3 和 12.8 点数），但选修这些课程的入学者较少；信息系统和物理学的差值则与数学接近（均为 10.7 点数）。这些比较无法证明课程结果是否先于专业变更，也无法证明二者是否存在因果关系。这正是接下来的两幅图追踪修读数学并更换专业的学生的原因。

图3展示了在整个2014年秋季至2019年秋季观察窗口内修读过 MAT 课程并更换专业的学生之间的专业变更流（包括新入学者和转学生）。图中的每条带状图元代表一定数量的学生，因此图中仅绘出涉及至少十名学生的流：共415条，涵盖19,489名学生中的17,164名（88.1%）。其余610条规模较小的流未予绘出。阅读环形图时：每个扇区代表一个专业，其大小取决于该专业在所绘流中转出和转入的学生总数；带状图元从转出专业指向转入专业，颜色与转出专业一致。每个扇区均标有对应专业的名称。图中凸显出三个结构性事实。首先，容纳未定专业和通识学习学生的大学学院是主要的转出方：大片绿色带状图元从该学院扇出，流向各个学院，这正是咨询单位应发挥的作用。其次，接收学生较多的专业分布在多个学院（生物学；运动机能学、健康与营养学；管理学；心理学；市场营销；计算机科学），因此，修读 MAT 课程学生的专业变更是全校范围内的流动。第三，已确定专业之间的流动主要集中在各自所属的学院内（带状图元往往停留在同一颜色区域附近），但也有明显的跨学院流动，流向商学和健康相关专业。表D1按字母顺序列出各专业，并标明每个扇区的角位置和大小（附录 D）。

3.5 专业变更者的相对学科评分

图4比较了每位学生在某一学科的成绩与其在所有其他学科的平均成绩。我们对参考队列中初始专业相同且有记录显示曾变更专业的学生（下文称为专业变更学生）的这些差值取平均。行表示学科，列表示学生的初始专业。负值表示这些学生在该行所示学科取得的成绩低于其他学科，正值则表示成绩较高。公式（A3）给出了计算方法。左侧的两列汇总了每一行中有观测值的单元格。第一列报告每位专业变更学生的平均差值（以平均绩点分数计）；第二列将该平均值乘以观测到的专业变更学生人数，因此也反映了选修该学科的专业变更学生人数。

这些学科分为两组。在图的上方一组中，专业变更学生的成绩低于他们在其他学科的成绩。按每位学生计算，化学的负差值最大（-0.50 个平均绩点分数；涉及14个初始专业的2,047名专业变更学生），其次是数学和政治学（各-0.43 个），再其次是商法、经济学和生物学（-0.39 至 -0.38个）。按专业变更学生总数计算，数学和政治学的负差值最大（分别为-1,662和 -1,593个平均绩点分数，涉及3,851名和3,708名专业变更学生）：这两门学科在全部30个初始专业列中都有观测值，且几乎所有初始专业对应的单元格都为负值。化学排名第三（-1,028个）。在图的下方一组中，专业变更学生的成绩高于他们在其他学科的成绩。按每位学生计算，信息系统（$+0.61$个）和艺术（$+0.57$个）领先；按总数计算，一年级研讨课 AIS 领先（涉及4,036名专业变更学生，共$+1$,487个，平均每位学生 $+0.37$ 个），几乎所有专业变更学生都会选修该课，其次是信息系统（$+925$个）。列聚类还显示了一个相关现象：离开工程、科学和商学专业的学生（右侧列组，从电气与计算机工程到金融学）的对比最为明显；在矩阵中，他们在上方各行的负值单元格和下方各行的正值单元格程度最强。成绩是对计入平均成绩的课程行取平均，因此退课和学分或无学分课程行不

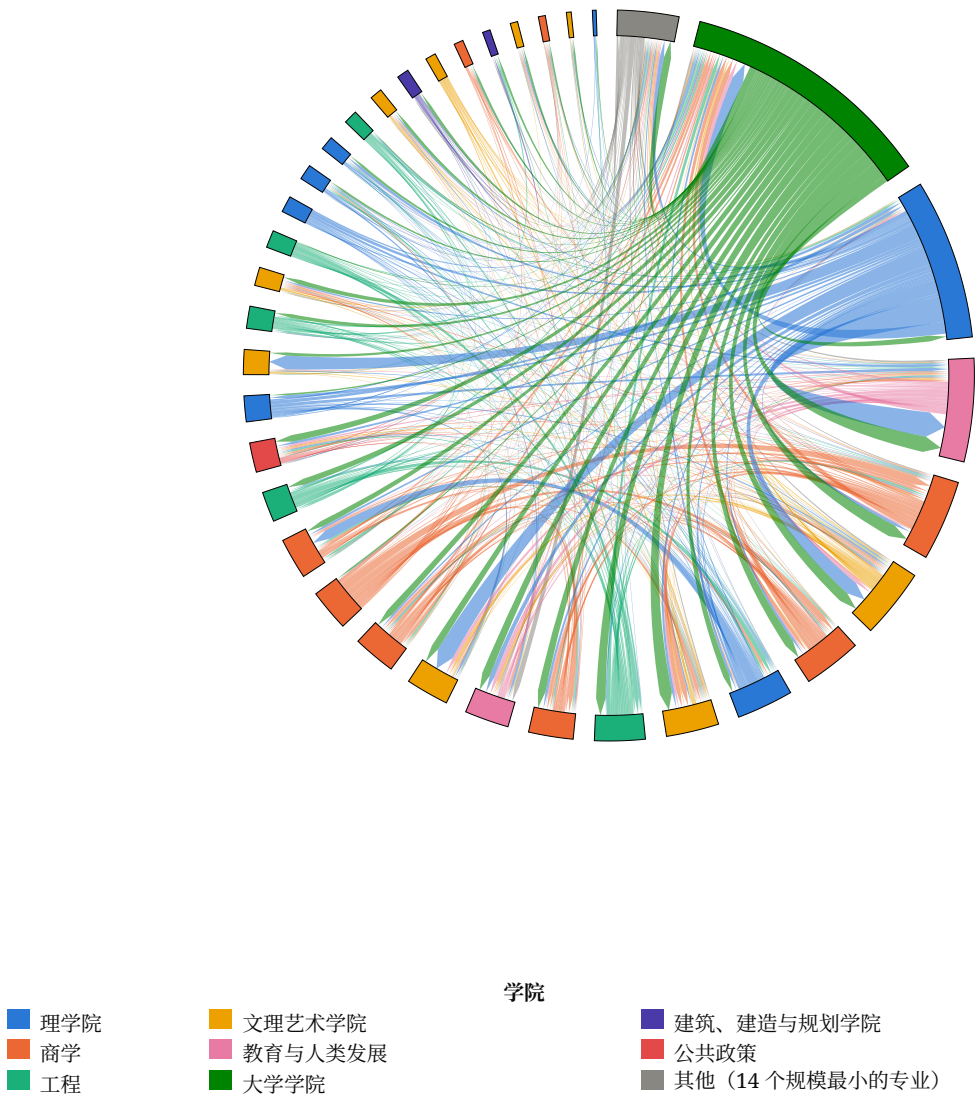


图3：修读过 MAT 课程并更换专业的学生之间的专业变更流（2014年秋季至2019年秋季）。图中绘出了至少涉及 10 名学生的 415 条流，共涵盖 19,489 名学生中的 17,164 名；较小的流未予绘出。每个扇区代表一个专业；弧长与该专业转出和转入的学生人数总和成正比。带状图元从转出专业指向转入专业，并沿用转出专业的颜色；颜色表示学院。名为“通识”的专业，是括号中所列学院面向全院学生的未定专业或通识专业。流量最小的 14 个专业归入标注为“其他”的灰色扇区。表D1列出了每个扇区的规模。

未进入。以上内容均是对专业变更学生成绩的描述；图中并未对难度与离校之间的因果关系作出排序。偏离项反映的是相对记录成绩。准备程度、课程构成、评分惯例以及转专业的选拔机制都可能造成这些偏离，因此成绩差距并不能衡量某一学科的教学要求有多高。未观测到的单元格在聚类时按零处理，因此行和列的排列顺序仅用于展示，树状图也不能证明存在自然形成的学术分组。

3.6 班级规模与成绩

图5绘出了班级平均GPA与班级规模的关系，每个学科对应一条平滑曲线，线宽与所服务的学生人数成正比。两点一目了然。不同学科的班级平均GPA在所观测的班级规模范围内总体较为平稳；相比班级规模，学科类别对曲线的区分作用要大得多：在任一固定规模下，不同学科之间的纵向差距约为两个GPA点。此外，在图示的班级规模范围内，MAT 曲线（红色）在学生人数较多的学科中始终最低，班级规模最小时平均GPA约为 2.4，在接近 200 名学生时降至约 1.8，此后趋于平缓；在某些特定规模下，少数学生人数较少的曲线会低于它。这些曲线比较的是同一学科不同课程的教学班，因此描述的是评分结果，并不构成设计上的对比。

预先设定的班级规模模型为首次入学学生提供了这一对比：其主要模型比较同一课程、同一学期内的入学者（第 2.4 节）。本小节陈述数学方面的三项结果。附录B.8完整报告了这些模型，包括每个学科的估计值（图B7）、情景（表B5）和敏感性检验。

对于入学数学成绩较低的入学者，数学课程班规模越大，*DFW*结果越多。班级规模与结果之间的关联取决于准备程度，正如 H4 所提出的，且结果支持 H4。入学数学成绩每增加一个标准差，班级规模的系数上升 0.046 个绩点（0.005 至 0.088；$p = 0.027$），DFW系数则下降 2.4 个百分点（0.8 至 4.0；$p = 0.007$）。给定入学分数时，班级规模与结果之间的关联是另一项对比，具有各自的区间。对于入学分数低于均值一个标准差的入学者，班级规模加倍与 DFW概率高出 3.3 个百分点（0.2 至 6.4；$p = 0.038$）以及成绩低 0.089 点相关（-0.183 至 $+0.005$；$p = 0.063$）。对于入学分数高于均值一个标准差的入学者，拟合得到的关联分别为 -1.5 个百分点（-3.7至 $+0.8$）和 $+0.003$ 个绩点（-0.072 至 $+0.079$）。因此，在入学分数较低的情况下，DFW关联与零有显著差异；该分数水平下的成绩关联方向相同，但其区间延伸至零。

对所有入学者取平均后，这种关联较弱。在所有数学课程班中，班级规模每加倍，系数为 -0.255绩点（-0.373 至 -0.138），呈现出下降曲线。加入课程与学期效应后，系数为 -0.026，因此总体关系在很大程度上反映了不同课程开设之间的差异：记录中微积分 I 和 II 的讲座班级中位数为36名学生，没有一个班超过50人；而微积分以下课程的教学班中，每六个就有一个超过100人。在加入学生协变量和教师效应的主要模型中，系数为 -0.043 （-0.117 至 $+0.031$；$p = 0.26$），DFW 比率也相应相差 $+1.0$ 个百分点（-1.3至 $+3.3$）。估计结果的方向与 H3 的预测一致（班级规模越大，数学成绩越低），且其区间包含零，因此未达到预先设定的 H3 判定标准。在纳入模型的20个学科中，汇总后的主要系数为每加倍 -0.002 绩点（-0.026 至 $+0.022$）；在数学以外的学科中，规模与成绩交互项的正负方向不一。

结果提出了一个可供检验的假设：缩小教学班规模。第 2.4 节的情景是将每个学生人数超过 45 人的数学讲座班拆分。对于首次尝试中成绩最低的三分之一

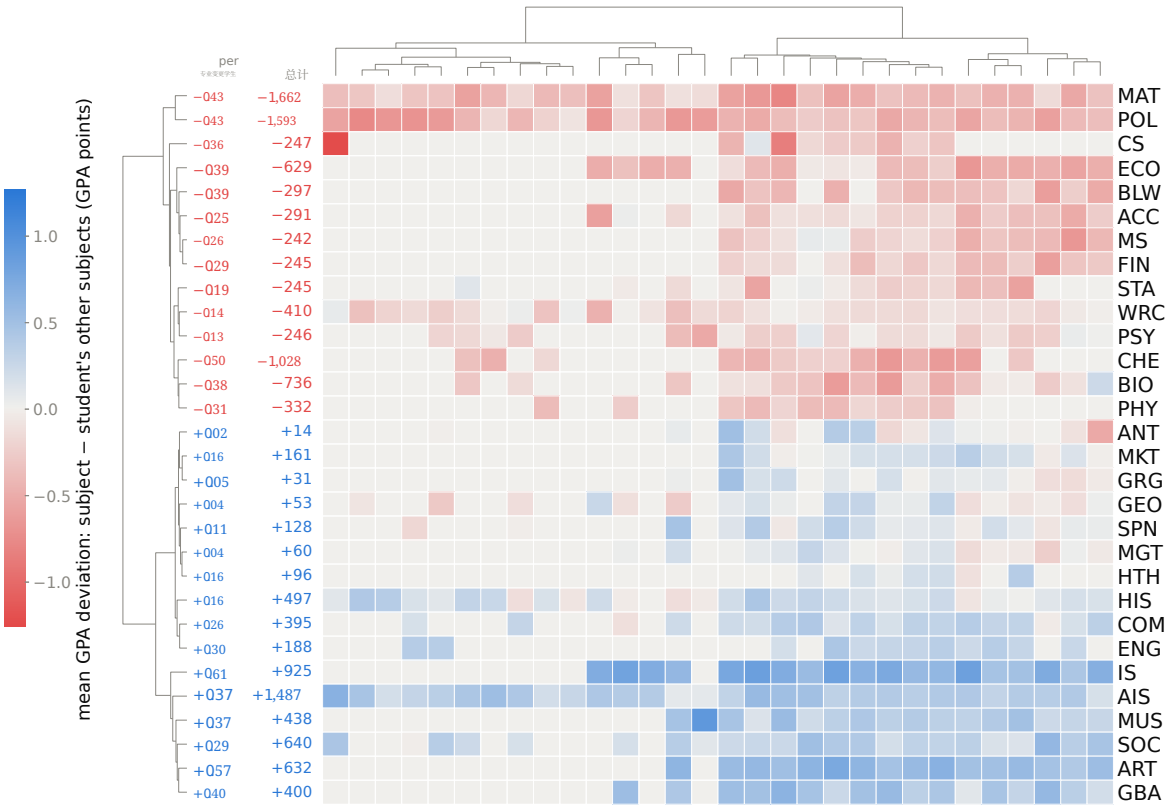


图4：更换专业学生中，某一学科的成绩与所有其他学科成绩之间的差异：参考队列中在观察窗口（第2.2节）内初始专业与最后专业不同的首次入学新生。行表示学科，列表示学生的初始专业。每个单元格表示初始专业为该列所示专业的学生在该行学科的平均绩点，减去他们在所有其他学科的平均绩点（公式（A3））。红色表示该学科成绩低于学生其他学科的成绩；蓝色表示成绩较高。成绩为计入平均成绩的课程行的平均值。少于15名学生的单元格留空；图中仅显示样本量最高的30个学科和初始专业。最左侧两列汇总各学科在其有观测值的单元格中的数据：平均差值（每名专业变更学生的 GPA 分数）以及该平均差值乘以观测到的专业变更学生人数（所有专业变更学生的 GPA 分数总计）。行和列按 Ward层次聚类排序；未观测到的单元格在聚类中按零值处理。

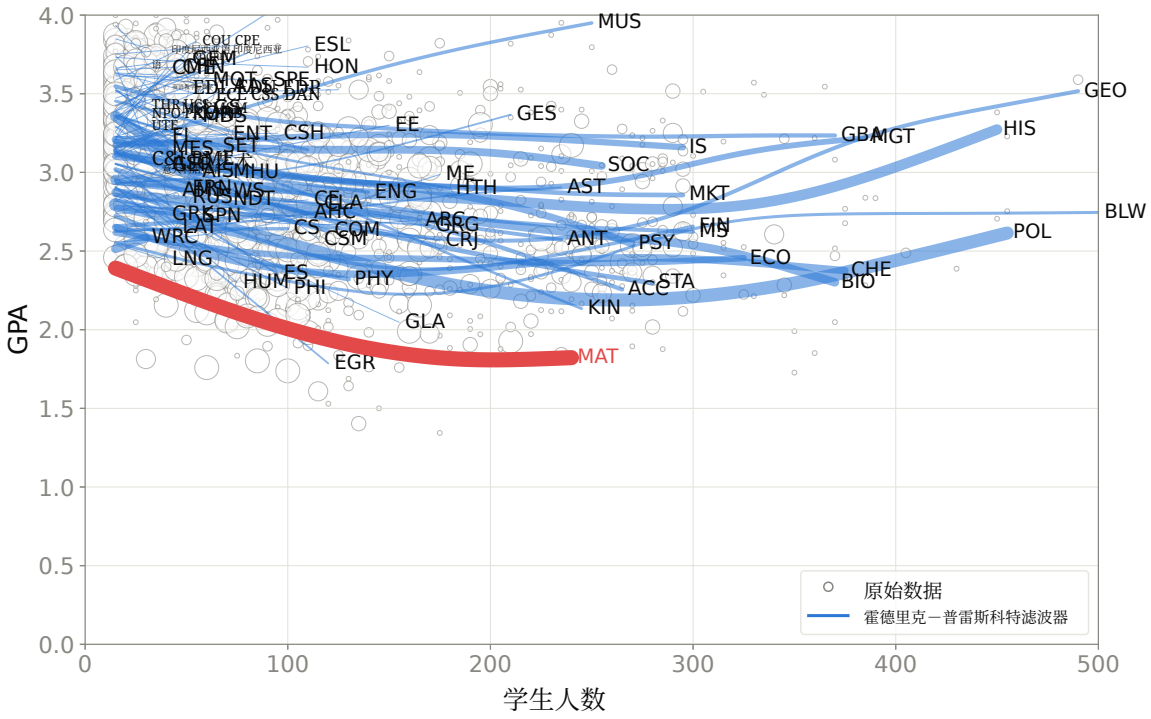


图5：2014年秋季至2019年秋季各观测学科的班级平均GPA（计入平均成绩的课程行）与班级规模的关系；图中网格显示至500名学生。每条曲线代表一个学科：班级规模按每5名学生为一个区间分组，每个区间内各观测规模的权重相同，经插值后使用Hodrick–Prescott滤波器平滑（$\lambda = 10000$；附录A.4）；线宽与所服务的学生人数成正比；MAT以红色绘制。每个气泡显示一个班级规模组的班级平均GPA。气泡面积与该组中不同班级规模的数量的平方成正比。只有学生人数超过10人的教学班才纳入分析。

入学成绩方面，预测结果为：每100次受影响的首次修读将减少4.6个DFW结果（-8.9 至-0.4；$p = 0.034$），并增加0.12个绩点（-0.01 至 $+0.25$；$p = 0.082$）。中间三分位的DFW结果预计减少1.6个（-5.1至 $+2.0$），最高三分位则预计增加2.0个（-1.3至 $+5.2$）。在所有受影响的首次修读中，预测变化为 -1.4 个百分点（-4.8 至 $+1.9$），而分班将使每年开设的教学班数量从246个增加111个。因此，预测的DFW结果降幅主要集中在准备程度较低的入学者中。只有在第 2.4 节所述条件下，这项预测才描述了一种干预；观察性模型并未证实这些条件。教学班由行政安排，因此本小节中的每项估计都反映一种关联，而结果所要求的检验应采用实验设计。

4 讨论

4.1 发现综合说明了什么

第 3 节的五项发现彼此相关。在全国范围内以及贝克萨尔县的学区中，儿童贫困程度越高，高中数学课程序列高阶部分的选修人数越少，而在贫困程度较低的地区则相反；

来自较贫困学区以及数学课程选修程度较低学校的入学者进入该院校时，数学成绩较低。入学成绩几乎与每个学科的第一学年成绩都存在关联，而第一学年未成功结果中，数学是最大的单一来源；出现这种结果的学生之后也未被观察到继续入学。每一条联系都是一种关联。从入学成绩开始的这些联系描述的是进入该院校的学生，而这些分析并未说明其中一条联系有多大程度上是由另一条联系造成的。

在不同设计下计算的同一院校记录的三种视角，将数学描述为一门评分偏低且选修人数广泛的学科。更换专业的学生在数学课程中取得的成绩低于他们在其他学科中的个人平均成绩；按每名学生计算，数学仅次于化学，而按总量计算则最高（图4）。在班级层面记录中，在所展示的班级规模范围内，MAT 是评分最低且选修人数众多的学科（图5）。在参考队列中，在观察窗口内任何时间修读 MAT 课程，都与在该窗口内没有高年级课程注册的较高几率相关；在使朴素 AIS 关联反转的队列限制下，这一估计仍然成立（表 B2）。这三种视角描述的是不同的人群和数量。它们构成了关联结果运行的背景；准备程度如何与表现相关的直接估计，则属于关联分析。

持续就读使这些视角具有实际意义。对最多入学者而言，数学是记录其艰难第一学年的课程：在那些之后未被观察到的入学者中，数学是不成功结果的最大单一来源（第 3.4 节）。因此，它是及早发现这些学生的实际切入点。该记录也限定了这一表述的范围。离校有许多记录无法涵盖的原因，而且记录无法区分转学与离开高等教育。在每一门入门学科中，DFW 都与之后未被观察到相关；而在其他学科没有 DFW 的数学入学者中，这一差异很小，其区间包含零。在仍然留校的入学者中，数学第一学年的 DFW 更常与记录中的专业变更同时出现，但具体幅度取决于哪些入学者仍可被观察到。这两种比较都无法厘清因果关系。

入学群体的发现为“入学机会型院校”这一说法赋予了可衡量的内涵：生源学区构成、生源学校记录的课程选修情况，以及入学者所持大学前学分的类型。该院校的生源来自贫困程度低于全州按儿童人口加权中心的学区（图 B2）；生源主要集中在两个大型城市学区，而这两个学区的微积分选课人数均略低于全国拟合线（图1）。学生入学时所持学分的类型与其学校记录的 AP 数学课程选修情况相对应：随着学校选修人数分档上升，持 AP 学分入学的学生比例上升，持双重注册学分入学的学生比例则下降（图 B1）；记录的 AP 数学课程选修人数为零的学校，则主要是规模最小的校区。这一入学群体描述将学生入学时持有的学分，与其生源学校和学区的背景结合起来；对许多学生而言，后者是在之后才测量的：学校概况采用的是 2017—18年CRDC数据快照，该快照的时间晚于许多学生的入学时间。这些概况描述的是与已入学学生总体相关的环境，并不能证明每名学生在上大学前实际获得了哪些机会。

这份方法学记录支持将结论推广到任何重复开展该项工作的院校。关联的方向可能取决于比较组：当比较范围限定为具有明确入学队列且时间跨度固定时，必修一年级研讨课的结果会从强烈不利转为强烈有利（附录 B.6）。同样，关联方向也可能取决于分母框架：在一万多个学区中，有七个框架不匹配地区会使按每名常住儿童计算的课程注册率的未加权全国梯度发生反转（附录 C.3）。目录中的每一项实质性来源缺陷都会在没有错误或警告的情况下产生看似合理的输出，而每一项缺陷最终都是通过针对其他问题的检查发现的。以下固定规则——队列限制、分母筛选条件、框架间比较检验，以及在暴露发生变化的层级计算不确定性——本身就是本文的结论。

全国排名构成了基线。在 2019–20 年的 Scorecard 数据中，该院校的完成率比根据其佩尔助学金占比拟合出的数值低 5.5 个百分点。在 2024–25 年数据中，在佩尔助学金占比基本不变的情况下（从 0.440 变为 0.441，见图 B5），其完成率比相应的拟合值高 6.4 点。这些是与各数据年份全国曲线进行的描述性比较。基线确定了第一个坐标，而这一变化跨越了动荡时期；这正是参照量存在的目的——用于衡量此类变化。

4.2 公平性与测量方面的启示

公平性视角从第3.3节测得的关联出发。通过分班测评工具、学期初信号和教学班层面的监测，入学准备程度或有助于识别需要学业支持的学生。学区和学校变量对预测准确率的贡献较小，但这并不能证明地理信息对于支持或公平性决策毫无必要：这一结论仅涉及两种评估设计下的这些预测变量，而且学区贫困程度在多个学科中的调整后系数仍为负值。这些数据也无法确定哪些学校或家庭过程造成了学生入学时已经存在的准备程度差异。

研究记录还指出一项院校能够掌控的因素：班级规模。第3.6节的推算具体到足以指导小班的设置方向。小班的益处集中在数学学科和准备程度较低的入学者身上；这些学生的入学成绩也与学区贫困程度和学校课程选修情况相关。相比之下，将每个大型数学课程班拆分都会使教学班数量增加45%，而拆分20个学科中的每个大型教学班则会使数量翻倍（表B5）。这是一个具有预先设定观察依据的假设。教学班由行政方式分配，因此需要通过设计分配方案，将准备程度较低的入学者安排到小班数学课程中进行检验。目前的比较无法证明，如果院校重新分配学生或增聘授课教师以拆分教学班，会产生什么结果。

记录的数学成绩与数学课程以外的表现也有关联。在定量学科之外，这种关联与语言成绩的关联程度相当。这两项成绩也反映了一般学业准备，而现有记录无法将这一因素与成绩本身区分开来（第4.3节），因此这些结果并不能证明数学支持会改善学生在其他学科的表现。

对不同课程类别的比较表明，学区层面的差异集中在数学课程序列的高阶阶段（第 3.2 节）。从入学者的角度看，也出现了相关模式：入学者在生源学校修读高等数学，与其入学分数以及第一门数学课程的层级有关联。全国成绩单证据表明，迈向学士学位的势头要到代数 II 之后才显现 [33, 第 30–33], 页，因此，与学区贫困程度呈现最强负相关的课程，正是根据 1982 年和 1992 年高中班级的资料，与学位完成关系最密切的课程。入学人数统计无法反映课程内容、修读时间及完成情况。在成绩单资料允许的情况下，网关测量程序应记录每位入学者在高中完成的最高数学课程。

招生政策和社区学院经费会影响哪些人进入大学，以及入学者带来多少学分。可选提交考试成绩的招生政策会改变有关申请者的信息，而第8号众议院法案重组了授予双重入学学分的社区学院经费 [35]。这些变化可能影响图B1所示的学分入学占比以及之后的结果。因此，2020年后比较应同时描述入学群体与表现，以免将观察到的差异自动归因于入学后的经历。

学区、学校和社区指标描述的是不同总体，只能通过经审计的交叉对照表进行关联（第2节）。在这些数据中，学区收入与高阶课程选课的关联比与广泛开设的核心课程选课的关联更强（第 3.2 节），因此，地区收入指标不能替代所有类型的课程选课指标。

专业变更结果揭示了入学人数快照无法反映的流动情况：学生会从大学学院转入各个学院的专业，而且可以确定转专业学生在哪些学科的成绩高于或低于自己的平均水平（图3和图4）。这些结果并不能说明某次转专业是否使特定学生受益。

4.3 局限性

本研究有九项局限性。其中一些源于研究仅涉及一所院校，并使用学校、学区和社区指标描述学生的背景。另一些则反映了相关信息缺失，例如哪些院校授予了双重注册学分，以及记录学生特征的时间。学校匹配的准确度也尚未经过验证，但可以通过审查现有材料进行评估。

1. **选定范围**。 每项与学生关联的统计数据都描述了某一所院校的学生（第1节）。这些结果不能推广至得克萨斯州的高中毕业生或全国范围。

2. **学校背景数据年份与链接准确性**。 附录 B.2 的生源学校学分概况使用 CRDC 2017–18 年快照，其时间晚于许多入学者的高中就读年份。关联模型采用最接近每位入学者高中毕业年级的周期，且时间相差不超过一年。2013–14 周期仅将双重注册记录为一项指标，因此，对于被分配至该周期的 8,180名关联入学者，双重注册率采用的是 2015–16 年的数据，最晚可能比其高中毕业年级晚两年。在这两种用途中，这些变量都描述了入学前后的学校背景，却未衡量每位学生自身的课程作业。在 2015–16 年与 2017–18年周期之间，人均比率只能不完全地反映生源学校的排序（秩相关为 0.62 至 0.78），因此，学校层面的解读不可避免地带有规模未知的测量误差。尚未通过经裁定的样本衡量已接受学校匹配的准确度，而关联模型使这一准确度问题更加重要；下一步应对已接受的匹配进行分层抽样审查，以审核链接情况。

3. **生态背景**。 机会地图集的结果测量指标具有双重生态层面特征（按 ZCTA 汇总的预测队列结果平均值），而图1中的收入中位数变量是随该地图集提供的美国社区调查 2012–16年普查区协变量。

4. **未观测到的学分合作方**。 这些记录没有双重注册学分的合作院校或授课地点字段，因此无法区分学分是在大学校园还是高中授课，也无法在不同授予学分的合作方之间固定评分标准。

5. **共享记录**。 在数学入门关卡上结论一致的三种研究设计共用同一份院校记录，因此也可能共同受到该记录缺陷的影响；这种一致性仅是在同一来源内的佐证，而在其他院校重复验证才是恰当的外部检验。

6. **关联性、经过选拔，且没有高中记录**。这些关联模型估计的是以入学为条件的关联。录取和入学会对总体进行筛选。入学选择可能改变背景与准备程度之间的关联，因此不应将这些关联推广到所有高中生。记录中没有高中平均成绩、排名或成绩单，因此两个入学分数反映了一般学业准备情况。数据摘录中包含一个助学金申请标记、一个调整后总收入（AGI）值，以及每名学生的一个最高

分数，但没有助学金年度和考试日期，因此这些变量相对于入学的时间关系，取决于数据摘录未说明的院校惯例；对于未申请助学金的14.4%学生，缺少 AGI；不纳入资助变量时，主要估计结果保持稳定。预先设定仅限于内部，且没有外部时间戳。方案冻结后作出了三项选择，并将其记录为偏离项：将范围限定为2014年秋季及之后的入学者、合并区间的联合协方差，以及仅使用训练数据构建预测目标。

7. **单一院校的持续就读**。 记录止于该院校。入学一年或更久后仍无课程记录的入学者，可能已离开高等教育或转学，因此此处的持续就读指在该院校持续注册。按第一年课程结果进行的比较未对准备程度和课程负荷作调整；这些比较是在分析计划定稿后才加入的，而且退课可能是已经开始的离校过程的一部分。

8. **缺乏研究设计的班级规模**。 班级规模模型比较行政分配下同一课程开课班级中的学生。研究未分析注册人数上限、教室和上课时间，因此未采用断点设计；同一课程的小班与大班之间仍可能存在差异。准备程度交互效应依据一项预先设定的检验（$p = 0.027$成绩为 0.007，DFW 为 0.007 ），仅针对数学，尚待重复验证。对于入学分数最低三分位的预测结果，自助法得出的区间显示，DFW结果不包含零，而成绩包含零；该情景也承袭了这些条件，以及其自身关于因果关系、线性和人员配置的假设。

9. **未按群体编制结果表**。本文没有按种族与族裔、佩尔助学金资格或第一代大学生身份列出队列结果，尽管入学群体的构成使这些维度对于一所西班牙裔服务型院校至关重要。关联模型按照预先设定的方案，并遵循附录 C.6 的披露控制制度（汇总单元格至少包含十名学生），将家庭资源、第一代大学生身份、性别和族裔纳入预测变量；下一步最直接的研究就是在该项目中编制按群体细分的结果表。

4.4 2020年后比较方案

基线支持采用一套具体的比较方案：对2020年秋季及之后入学的学生应用相同的分析流程，使用相同的观察窗口，并在总体定义、研究设计和生成表格均一致的条件下，将各项参照量与其基线对应量进行比较。重复这些分析可以显示哪些测量值发生了变化。如果测得的入学情况保持稳定，而学业进展相关指标发生变化，那么这种差异是在入学后出现的；但这并不能确定其原因。学分入学占比的变化反映入学群体的情况。成绩偏差或班级规模曲线的变化，可能反映学生、课程或评分实践有所不同。关联模型中的系数变化，则反映记录在案的准备程度与一年级成绩之间的关联有所不同。只要报告单位和定义在各数据年份间保持一致，Scorecard面板便可提供全国性比较。首个后续时期的坐标数据已经可用（附录B.5），而院校内部的比较将说明这些指标中有哪些在此期间发生了变化。

有三项变化使得与2020年前基线的比较更加复杂。紧急评分政策改变了所有基于GPA的工具的含义，因此，在比较2020和2021学期的偏差矩阵和班级规模时，必须将通过/不通过制占主导的学期视为单独的标记制度。可选提交考试成绩的招生政策减少了基线审计用于

选拔检查的标准化分数字段，因此后期审计需要采用替代协变量。此外，生成式人工智能在课程作业中的普及，可能分别对成绩和学习产生影响。基于成绩的测量指标描述的是记录成绩，而记录成绩的变化未必意味着学习也发生了相应变化。每种风险都会改变差异的含义，其中一些风险（包括2025年的合并）甚至可能直接破坏实体可比性；标明人群、数据版本和制度，是在仍可进行比较时保留每项比较之解释的做法。

总而言之，我们记录了某所广泛开放入学的西班牙裔服务型研究大学中，学区儿童贫困与课程选课、入学时的数学准备程度以及第一年表现之间的关联，并提供了2020年前的参照数量、关联基础设施、验证记录和估计目标定义。这些内容是开展2020年后比较所必需的，而2020—2025年的种种中断使此类比较成为必要。

声明

资助。 本材料基于美国国家科学基金会第 2518973 [48]号奖项资助的研究工作。本文所表达的任何观点、发现、结论或建议均属于作者本人，不一定反映美国国家科学基金会的观点。

竞争性利益。 作者声明不存在任何竞争性财务或非财务利益。

伦理审批。 对机构学生记录的分析依据机构审查委员会方案 FY24-25-316“通过人工智能增强教育支持学生在本科早期数学课程中的学习与成功”开展。该方案允许进行回顾性数据分析：对2014年秋季至2019年秋季期间就读学生的现有机构行政记录进行非干预性二次分析，不实施干预、不与学生联系或互动，也不收集新数据；所有暴露和结果窗口均在2020年春季教学中断之前结束。这些记录由院校根据数据申请 201908048-IN 提供，该申请于2019年8月25日获批，并授权将数据存储以供未来研究使用。披露控制制度见附录 C.6：个人记录在原处查询，绝不导入分析数据库；仅发布至少有十名学生支持的汇总单元格，并在已发布计数旁记录被抑制单元格的数量。

参与同意。 本研究未与学生接触或互动，也未收集任何新数据。这是对现有行政记录进行的二次分析，依据上述方案和数据申请开展。

数据可用性。公共来源数据（SAIPE、CRDC、美国国家教育统计中心 EDGE 和共同核心数据（CCD）、机会地图集、College Scorecard 以及人口普查关系文件）可从参考文献所列网址向其发布机构获取。随代码提供的清单列出了每个公开输入文件的数据集、版本年份、SHA-256 哈希值、大小和来源地址（共 378 个文件，均已与磁盘上的文件逐一核验）；另一份清单则记录了本研究获取的六个档案的获取情况。每张图和每个表均随附汇总输入数据，并由另一份清单列出。对于大多数图表，这些输入数据包含所展示的数值；而对于热图、班级规模曲线和队列表，其中包含的是汇总统计数据，重新绘制这些图表

则需要院校记录。根据方案和数据申请的条款，机构数据摘录受到限制，无法共享；获授权的分析人员可以通过计算复现其结果。

代码可用性。 本文的每个图、表和计算数值，均由分析代码仓库中的指定程序生成；该代码仓库随论文一并提供，以供审阅和发布。主脚本按照依赖顺序运行各程序；每项分析都会将结果写入统计文件，正文则按名称从这些文件中读取每个计算数值（附录 A.1）。关联入学者模型的分析计划及其章节规模补充文件均在估计前冻结，并与代码、哈希记录以及每次冻结后所作偏离的带日期清单一并提供。补编记录中列出的每个文件，均以其冻结状态收录。原始方案记录中列出的文件，凡自冻结以来未发生变化者，均以该状态收录。方案、数据组装程序和运行脚本在首次记录的偏离发生后进行了修订，随后还扩展了公开课程数据加载器；代码仓库包含它们的当前版本，而记录中的哈希值则标识其冻结状态。

作者贡献。 本研究由唯一作者设计，分析由其完成，手稿亦由其撰写。

生成式人工智能的使用。 作者以一份此前撰写、约 10,000 页的内部报告为依据，准备了初稿。准备手稿期间，作者使用了 Visual Studio Code 扩展，调用 Anthropic 的 Claude Code（Claude Fable 5.2 Max Ultracode）和 OpenAI 的 Codex（GPT-6.1 Sol Extra High）。这些 AI 工具协助修改文稿、综合分析结果，并生成 Python 代码，以取代现有的 MATLAB 分析流程。每个系统的输出都由另一个系统和作者进行了严格审查。审查分轮进行，分别检查整篇手稿、各个教学班和软件组件。每个组件至少经过九轮审查，旨在发现错误、缺乏依据的论断以及解释上的不足。未向任何一个 AI 系统提供个人层面的学生记录。定量分析通过代码在本地执行；代码查询存储在加密硬盘上的数据库，所用计算机在启动时及访问操作系统时均须进行身份验证。作者设计了研究、指导了软件开发，并审查和编辑了所有生成的内容。作者对研究的设计与实施、分析以及手稿内容承担全部责任。

参考文献

[1] 美国人口普查局。《2025年版城市和城镇人口估计》，2026年。网址：https: // www.census.gov/newsroom/press-releases/2026/vintage-2025-city-town-pop-est i mates.html。2025年7月1日人口估计。访问日期：2026年8月10日。[2] Craig Benson。《各州和大都市区贫困状况：2022年》。美国社区调查简报 ACSBR-016，美国人口普查局，2023年。网址：https://www.census.gov/content/dam /Census/library/publications/2023/acs/acsbr-016.pdf。访问日期：2024年12月13日。[3] 美国最高法院。《圣安东尼奥独立学区诉罗德里格斯案》，《美国判例汇编》第411卷第1页，1973年。网址：https://www.loc.gov/item/usrep411001/。《美国判例汇编》第411卷，1972年10月开庭期。访问日期：2026年8月10日。

[4] Anthony P. Carnevale、Ban Cheah 和 Emma Wenzinger。《大学回报：受教育程度提高并不总意味着收入增加》。技术报告，乔治城大学教育与劳动力中心，2021年。网址：https://cew.georgetown.edu/cew-reports /collegepayoff2021/。访问日期：2026年8月10日。[5] 美国劳工统计局。《教育带来回报：按教育程度划分的失业率和收入，2025年》，2026年。网址：https://www.bls.gov/emp/tables/unemploymen t-earnings-education.htm。就业预测项目，表5.1；数据年份为2025年，最后修改于2026年9月1日。访问日期：2026年11月20日。[6] Raj Chetty、John N. Friedman、Emmanuel Saez、Nicholas Turner 和 Danny Yagan。《美国各高校的收入隔离与代际流动性》。《经济学季刊》，135(3):1567–1633，2020年。doi：10.1093/qje/qjaa005。网址：https: //doi.org/10.1093/qje/qjaa005。[7] 美国教育委员会和卡内基基金会。《2025年卡内基分类：研究活动类别，得克萨斯州高校》，2025年。网址：https://carnegieclassifica tions.acenet.edu/institutions/?research2025%5B%5D=1&stabbr%5B%5D=TX。访问日期：2026年8月10日。[8] 得克萨斯大学圣安东尼奥分校。《南方学院与学校协会学院委员会批准合并 UTSA 和 UT Health San Antonio》，2025年。网址：https://news.utsa.edu/2025/06/sacscoc-approves-merger-of-utsa-and-ut-health-san-antonio/。2025年6月公告；合并于2025年9月1日生效。访问日期：2026年8月10日。[9] 得克萨斯大学圣安东尼奥分校机构研究与分析办公室。《UTSA事实手册：本科招生，2024年秋季》，2024年。网址：https://www.utsa.edu /ir/docs/resources/factbook/2024/UndergraduateAdmissions.pdf。2024年秋季首次入学的本科生：25,422人申请，22,063人获录取（87%）。访问日期：2026年8月10日。[10] 得克萨斯大学圣安东尼奥分校。《UTSA快速事实 2024–2025：入学人数》，2024年。网址：https://www.utsa.edu/_files/about/pdfs/fast-facts-enrollment.pdf。2024年秋季数据：总注册人数35,770人，60%为西语裔，43%为第一代大学生。访问日期：2026年11月20日（存档副本，SHA-256前缀 1970dc4f441a2d7e）。[11] 美国政府问责局。《K-12教育：贫困学生较多且规模较小的公立高中提供的大学预备学术课程较少》。技术报告 GAO-19-8，美国政府问责局，2018年10月。网址：https://www.gao.gov/products/gao-19-8。[12] Dylan Conger、Mark C. Long 和 Patrice Iatarola。《解析高级课程选修中的种族、贫困和性别差异》。《政策分析与管理杂志》，28(4): 555–576，2009年。doi：10.1002/pam.20455。网址：https://doi.org/10.1002/pam.20455。[13] Kristin Klopfenstein。《大学先修课程：少数族裔是否享有平等机会？》《教育经济学评论》，23(2):115–131，2004年。网址：https://ideas.repec.org/a/eee/ ecoedu/v23y2004i2p115-131.html。[14] Di Xu、Sabrina Solanki 和 John Fink。《全民大学加速学习？绘制大学先修课程和双重注册参与中的种族差距》。《美国教育研究杂志》，58(5):954–992，2021年。doi：10.3102/0002831221991138。网址：https://doi.org/10.3 102/0002831221991138。

[15] Mark C. Long、Dylan Conger 和 Patrice Iatarola。高中课程修读对中学及高等教育阶段成功的影响。《美国教育研究杂志》，49(2):285–322，2012年。doi: 10.3102/0002831211431952。URL https://doi.org/10.3102/0002831211431952。[16] Julian R. Betts 和 Darlene Morell。本科平均绩点的决定因素：家庭背景、高中资源和同辈群体效应的相对重要性。《人力资源杂志》，34(2):268–293，1999年。URL https://www.jstor.org/stable/146346。[17] Jason M. Fletcher 和 Marta Tienda。大学学业成就中的种族与族裔差异：就读的高中是否重要？《美国政治与社会科学学会年鉴》，627(1):144–166，2010年。doi: 10.1177/0002716209348749。URL https: //pmc.ncbi.nlm.nih.gov/articles/PMC3489015/。[18] Sandra E. Black、Jane Lincove、Jennifer Cullinane 和 Rachel Veron。你能把高中经历抛在身后吗？《教育经济学评论》，46:52–63，2015年。doi: 10.1016/j.econedurev.2015.02.003。URL https://doi.org/10.1016/j.econedurev.2015.02.003。[19] Gregory C. Wolniak 和 Mark E. Engberg。大学第一学年的学业成就：高中背景普遍影响的证据。《高等教育研究》，51(5):451–467，2010年。doi: 10.1007/s11162-010-9165-4。URL https://doi.org/10.1007/s1 1162-010-9165-4。[20] Elaine M. Allensworth 和 Kallie Clark。高中平均绩点和 ACT成绩作为大学毕业的预测变量：考察各高中之间一致性的相关假设。《教育研究者》，49(3):198–211，2020年。doi: 10.3102/0013189X20902110。URL https://doi.org/10.3102/0013189X20902110。[21] Michael N. Bastedo、Mark Umbricht、Emma Bausch、Bo-Kyung Byun 和 Yiping Bai。情境化的高中表现：为公平的综合评估、免考试和免测试招生政策提供依据的证据。*AERA Open*，9，2023年。doi: 10.1177/23328584231197413。URL https://doi.org/10.1177/23328584231197413。[22] Leonard Ramist、Charles Lewis 和 Laura McCamley-Jenkins。在预测大学成绩时的学生群体差异：性别、语言和族裔群体。技术报告，College Board报告第93-1号；ETS RR第94-27号，美国大学理事会，1994年。URL https://files.eric.ed.gov/fulltext/ED562876.pdf。[23] Jesse M. Rothstein。大学表现预测与 SAT。《计量经济学杂志》，121(1–2):297–317，2004年。doi: 10.1016/j.jeconom.2003.10.003。URL https://doi.org/10.101 6/j.jeconom.2003.10.003。[24] Todd R. Stinebrickner 和 Ralph Stinebrickner。了解学术能力与大学辍学决策。《劳动经济学杂志》，30(4):707–748，2012年。doi: 10.1086/66 6525。URL https://doi.org/10.1086/666525。[25] Carmen Astorne-Figari 和 Jamin D. Speer。专业变更是重大的变更吗？成绩、性别和偏好在大学专业转换中的作用。《教育经济学评论》，70:75–93，2019年。doi: 10.1016/j.econedurev.2019.03.005。URL https://doi.org/10.1016/j.econedurev.2019.03.005。

[26] Ralph Stinebrickner 和 Todd R. Stinebrickner。理科专业？大学专业与辍学的初始信念和最终结果。《经济研究评论》，81(1):426–472，2014年。doi: 10.1093/restud/rdt025。URL https://doi.org/10.1093/restud/rdt025。[27] Edward C. Kokkelenberg、Michael Dillon 和 Sean M. Christy。班级规模对公立大学学生成绩的影响。《教育经济学评论》，27(2):221–233，2008年。URL https://ideas.repec.org/a/eee/ecoedu/v27y2008i2p221-233.html。[28] Maria De Paola、Michela Ponzo 和 Vincenzo Scoppa。班级规模效应对学生学业成就的影响：不同能力和学科领域之间的异质性。教育经济学，21(2):135–153，2013年。doi: 10.1080/09645292.2010.511811。URL https://doi.org/10.1080/09645292.2010.5118 11。[29] Benjamin N. Matta、Joseph M. Guzman、Sue K. Stockly 和 Benjamin Widner。西班牙裔服务型院校中的班级规模效应与学生表现。黑人政治经济学评论，42(4):443–457，2015年。doi: 10.1007/s12114-015-9214-5。URL https://doi.org/10 .1007/s12114-015-9214-5。[30] 美国教育部、民权办公室。美国公立学校数学、科学和计算机科学课程及学术课程中的学生参与机会与入学人数：2020–21年民权数据收集。技术报告，美国教育部，2024年5月。URL https://www.ed.gov/media/document/crdc-student-acces s-enrollmentpdf-21419.pdf。访问日期：2024年11月3日。[31] 得克萨斯州议会。第5号众议院法案，得克萨斯州第83届议会常规会议，2013年。URL https: //capitol.texas.gov/billlookup/text.aspx?LegSess=83R&Bill=HB5。访问日期：2025年1月2日。[32] 得克萨斯州教育局。数学常见问题，2024年。URL https://tea.texas.gov/educators/subject-areas/mathematics-frequently-asked-questions。基础高中课程计划中的代数 II要求。访问日期：2026年11月20日。[33] Clifford Adelman。重访工具箱：从高中到大学完成学位的路径。技术报告，美国教育部，华盛顿特区，2006年2月。URL https://files.eric.ed.gov/fulltext/ED490195.pdf。ERIC ED490195；高中阶段最高数学课程与学士学位取得，第30–33页。访问日期：2026年11月20日。[34] Nicholas W. Hillman。大学机会地理：教育荒漠案例。《美国教育研究杂志》，53(4):987–1021，2016年。doi: 10.3102/00028312166532 04。URL https://doi.org/10.3102/0002831216653204。[35] 得克萨斯州议会。第8号众议院法案，得克萨斯州第88届议会常规会议：社区学院经费，2023年。URL https://capitol.texas.gov/billlookup/History.aspx?LegSess=88R&Bill=HB8。访问日期：2026年8月10日。[36] 数字教育委员会。数字教育委员会2024年全球人工智能学生调查，2024年。URL https://www.digitaleducationcouncil.com/newsroom/what-students-want-key - results-from-dec-global-ai-student-survey-2024。对16个国家3,800多名学生开展的国际调查。访问日期：2026年8月10日。

[37] Heather L. Schwartz 和 Melissa Kay Diliberti。《更多学生使用人工智能做作业，更多学生认为这会损害批判性思维：美国青年调查组的部分发现》。研究报告 RR-A4742-1，兰德公司，美国加利福尼亚州圣莫尼卡，2026年。网址：https: //www.rand.org/pubs/research_reports/RRA4742-1.html。2025年12月对1,214名在学、年龄为12–29岁的青少年开展的调查。访问日期：2026年8月10日。[38] Raj Chetty、John N. Friedman、Nathaniel Hendren、Maggie R. Jones 和 Sonya R. Porter。《机会地图集：社会流动性的童年根源》。工作论文 25147，美国国家经济研究局，2018年。网址：https://www.nber.org/papers/w25147。访问日期：2025年1月3日。[39] A. Colin Cameron、Jonah B. Gelbach 和 Douglas L. Miller。《基于自助法的改进：聚类误差下的推断》。《经济学与统计学评论》，90(3):414–427，2008年。doi: 10.1162/rest.90.3.414。网址：https://doi.org/10.1162/rest.90.3.414。[40] Oriana Bandiera、Valentino Larcinese 和 Imran Rasul。《班级规模的异质性效应：来自大学生面板数据的新证据》。《经济学杂志》，120(549):1365– 1398，2010年。doi：10.1111/j.1468-0297.2010.02364.x。网址：https://doi.org/10.1111/j.1468 -0297.2010.02364.x。[41] Alexei Karas。《班级规模对成绩和课程评价的影响：来自多班课程的证据》。《经济研究公报》，73(4):624–642，2021年。doi: 10.1111/boer .12274。网址：https://doi.org/10.1111/boer.12274。[42] 美国国家教育统计中心。教育人口与地理估算（EDGE）：学区地理关系文件，2017年，未注明日期。网址：https://nces.e d.gov/programs/edge/Geographic/RelationshipFiles。普查区与学区关系文件（GRF17）：TIGER 2017学区边界，代表2015–16学年，基于2010年人口普查区；土地面积份额。[43] Juan B. Guti´errez。《UTSA学生成功的相关因素》。校内机构报告，德克萨斯大学圣安东尼奥分校数学系，美国得克萨斯州圣安东尼奥，2019年。网址：https://mathresearch.utsa.edu/。十卷本机构报告；须经院校批准方可查阅。[44] Richelle M. Blair、Ellen E. Kirkman 和 James W. Maxwell。《美国数学科学本科项目统计摘要：2015年CBMS调查》。美国数学学会，美国罗得岛州普罗维登斯，2018年。ISBN 9781470442514。网址：https://www.ams.org/profession/data/cbms-survey/cbms2015。按课程层级划分的入学人数：表S.2。访问日期：2026年8月10日。[45] Xianglei Chen。《美国公立两年制和四年制院校的补习课程修读：范围、经历与结果》。统计分析报告 NCES 2016-405，美国教育部、美国国家教育统计中心，2016年。网址：https://nces.ed.gov/pu bs2016/2016405.pdf。访问日期：2026年8月11日。[46] Thomas Bailey、Dong Wook Jeong 和 Sung-Woo Cho。《社区学院发展性教育课程序列中的转介、入学与完成情况》。《经济学与教育评论》，29(2):255–270，2010年。doi: 10.1016/j.econedurev.2009.09.002。网址：https://doi.org/10.1016/j.econedurev.2009.09.002。

[47] Alexandra W. Logue、Mari Watanabe-Rose 和 Daniel Douglas。《被评估为需要补习数学的学生是否应改修大学层次定量课程？一项随机对照试验》。《教育评估与政策分析》，38(3):578–598，2016年。doi：10.3102/0162373716649056。网址：https://doi.org/10.3102/0162373716649056。[48] 美国国家科学基金会。资助项目摘要 2518973，2025年。网址：https://www.nsf.gov/ awardsearch/showAward?AWD_ID =2518973。访问日期：2026年8月10日。[49] 美国人口普查局。2010年邮政编码制表区至普查区关系文件，未注明日期。网址：https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/2010-zcta-record-layout.html。文件 zcta tract rel 10.txt；基于2010年人口普查区定义的人口重叠权重，与机会地图集的数据版本相匹配。[50] Yair Mundlak。《关于时间序列与横截面数据的合并》。《计量经济学》，46(1): 69–85，1978年。doi：10.2307/1913646。网址：https://doi.org/10.2307/1913646。[51] Yoav Benjamini 和 Yosef Hochberg。《控制假发现率：一种实用而有效的多重检验方法》。《皇家统计学会会刊：B辑（方法论）》，57(1):289–300，1995年。doi：10.1111/j.2517-6161.1995.tb02031.x。网址：https://doi.org/10.1111/j.2517-6161.1995.tb02031.x。[52] Rebecca DerSimonian 和 Nan Laird。《临床试验中的荟萃分析》。《受控临床试验》，7(3):177–188，1986年。doi：10.1016/0197-2456(86)90046-2。网址：https://doi.org/10.1 016/0197-2456(86)90046-2。[53] A. Colin Cameron、Jonah B. Gelbach 和 Douglas L. Miller。《多维聚类下的稳健推断》。《商业与经济统计学杂志》，29(2):238–249，2011年。doi：10.1198/jb es.2010.07136。网址：https://doi.org/10.1198/jbes.2010.07136。[54] 美国教育部。College Scorecard原始数据，2026年6月发布版，2026年。网址：https://collegescorecard.ed.gov/data/。数据年份为2019–20和2024–25；存档文件 College Scorecard Raw Data 06102026.zip，SHA-256 前缀 bfe3add63069fea1；参见论文0数据清单。[55] 美国人口普查局。小区域收入与贫困估计（SAIPE）：学区估计，1995–2024，2026年。网址：https://www.census.gov/data/datasets/2024/demo/s aipe/2024-school-districts.html。收入年份1995至2024年的估计（28个数据版本）；2024年的估计于2026年2月发布。批量学区文件在导入时附有SHA-256清单；每项分析均注明所用数据版本。[56] 美国教育部民权办公室。民权数据收集（CRDC）：公共使用学校文件，2011–12至2021–22，未注明日期。网址：https://civilrightsdata.ed .gov/。文件来自2011–12、2013–14、2015–16、2017–18和2021–22各轮收集，每轮均单独发布；所用文件及其SHA-256哈希值列于代码包的源文件清单中。[57] 美国国家教育统计中心。教育人口与地理估算（EDGE）：公立学校位置，未注明日期。网址：https://nces.ed.gov/programs/edge/Geogra phic/SchoolLocations。学校点位位置数据，包含2017–18和2021–22两个数据版本，各自单独发布；所用文件及其SHA-256哈希值列于代码包的源文件清单中。

[58] 美国国家教育统计中心。共同核心数据（CCD）：2015 年公立小学/中学学校总体调查。网址：https://nces.ed.gov/ccd/pubschuniv.asp。用于恢复目录的 2010–11 和 2013–14 学校总体文件；这些文件记录了学校所在城市和年级范围。

A 定量方法

本附录首先介绍本研究在严谨性与可重复性方面的记录，随后列出各估计量及其对应的图或表（如有）。每项分析均基于各自的总体和观察窗口进行计算，详见表 C1。附录 C.4 介绍了适用于这些估计量的推断惯例，代码可用性声明则列出了实现脚本。门户课程与学业进展之间的关联分析采用具有固定观察窗口的参考入学队列；附录 B.6 报告了朴素比较/队列比较，该比较说明了为何应在计算关联之前先定义入学人群。

A.1 严谨性与可重复性

代码仓库。本研究的所有程序都存放在同一个代码仓库中，该仓库随论文一并提供，并公开托管于 https://github.com/biomathematicus/gateway。仓库根目录包含仓库说明、主脚本、公开输入文件清单以及预先设定的分析记录。一个文件夹存放程序，另一个文件夹存放本手稿读取的汇总输出。主脚本按依赖顺序运行这些程序：载入公开数据、开展学区分析、将高中名称与学校匹配、确定机构基线、运行关联模型和班级规模模型，以及生成本手稿中的数值。图和表格由计算统计量的同一批程序导出。

手稿中没有任何计算所得的数值是手动输入的。每项分析都会将结果写入与其图对应的统计文件。宏模块为文本中打印的每个数值命名，并从该文件或设定该数值的程序常量中读取其值。随后，分析流程会为每项分析写入两个 LaTeX 文件。第一个文件使用 `\newcommand` 声明每个数值，并将它们的初始值全部设为同一条红色警告信息：`[RUN THE PIPELINE]`。第二个文件则使用 `\renewcommand`为每个计算所得的数值赋值。对于微积分选课人数的全国平均值（第 3.1 节），这两个文件包含以下几行：

```
\newcommand{\CalcMeanPerHundred}{\WarningRunPipeline}
\renewcommand{\CalcMeanPerHundred}{4.6}
```

手稿在导言区加载声明，并在文档开始后加载数值。源文件中写的是 \CalcMeanPerHundred，读者看到的则是 4.6。因此，任何程序都未提供数值的量，无论用于正文、表还是图注，都会显示为红色警告，让分析流程缺失或未完整运行的情况一目了然。十个宏模块为 856 个量命名。检查程序会列出源文件中仍然直接写出的每个数字，并将其与保留为文本的字面量清单进行比对，同时注明各自的原因：日历标签、包含数字的名称、分析单位、公式中的数字、引用来源中引述的统计数据，以及统计惯例。

输入与已发布输出。上下文数据库由源文件重新构建，并记录其哈希值。代码仓库中的清单列出了每个公开输入文件及其哈希值（共 378 个文件，均已与磁盘上的文件逐一核验），另有一份清单列明每张

图和每个表对应的程序及已发布输入。已发布输出是一份明确列出的汇总文件清单，每个文件均按照附录 C.6 的披露规则进行扫描。

受限记录。机构数据摘录受到限制，因此获授权的分析人员可以通过计算重现其结果，而代码仓库则保存了手稿所读取的汇总输出。较早的机构基线程序通过本地查询缓存读取该摘录。每个缓存文件都与生成它的查询的哈希值一同存储；只要查询与记录的不一致，程序就会重新读取源数据。在本次提交的运行中，这些程序是在空缓存上执行的，它们写出的每张表、每个统计文件和每个底层数据文件都与先前的输出完全一致。哈希值标识查询；若源表本身发生变化，则需使用代码中说明的刷新选项。加载程序会应用既定的数据规则（附录 C.3），以确保其中记录的缺陷不会通过未来的分析代码悄然卷土重来。

预先设定的分析。关联入学者模型和班级规模模型均在估计前设定，且分析流程首先在打乱后的结果上端到端运行（第 2.4 节）。方案、补编、它们的哈希记录以及带日期的偏离事项清单均位于代码仓库中；代码可用性声明说明了仓库保存了哪些冻结文件。

A.2 记录关联

记录关联评分。 高中名称通过优先包含匹配评分进行解析。设 A和 B 为去除学校类型词元后两个名称的词元集合。当包含度$c(A,B) = |A \cap B| / \min(|A|, |B|)$、杰卡德相似度 $J(A,B) = |A \cap B| / |A \cup B|$以及字符三元组相似度 $t(A,B)$ 时，匹配评分为

$$s(A,B) \;=\; 0.55\, c(A,B) \;+\; 0.20\, J(A,B) \;+\; 0.25\, t(A,B), \tag{A1}$$

在城市、邮政编码、标准化城市键和州构成的分层区块内计算。自动接受要求满足 $s \geq 0.72$；在州级区块内，还要求比次优匹配项至少高 0.08，而在城市和邮政编码区块内，区块本身即可完成消歧，因此不要求满足差值条件（附录 C.2）。

A.3 学区课程梯度

加权最小二乘法与诊断。 学区课程梯度是课程成绩中每名估算学生对应的选课学生人数对儿童贫困率的最小二乘斜率，并以该课程的入学人数作为权重；标准误按州聚类，在仅包含一个州的数据中拟合时，则使用异方差稳健标准误（图1）。分析范围包括匹配后的学区中 5–17 岁儿童至少有 200 人（采用附录 C.3 的处理前分母规则），且报告了代数 II 选课人数的学区；这一指标在此用于标记运营高中课程序列的学区。作为稳健性变体，另采用中性框架，纳入所有报告了任何高中核心数学或科学课程的学区。

不同课程类别间的相对梯度。 在该共同样本框内，将加权估计量应用于民权数据收集（CRDC）学校文件中报告的十五个课程类别：7至8年级、9至10年级和11至12年级的代数 I；几何；代数 II；高等数学；微积分；生物学；化学；物理学；任意课程、数学课程和科学课程中的大学先修课程（AP）课程选修人数；国际文凭课程；以及双重

注册。在提供高中成绩的学区内，课程项目中的民权数据收集（CRDC）跳过代码表示学校报告该学科的班级数为零，并将其作为零计入学区总数。其他所有保留代码均表示数据缺失；如果某个学区在任何课程类别中出现一个此类代码，就会被排除在共同样本框之外（2017–18年为9,795个学区中的5个），因此所有十五个课程类别均在相同的9,790个学区上进行比较（附录 C.3）。原始斜率会随每门课程的基准比率而变化，因此课程 c 的报告量是相对梯度

$$g_c \;=\; 100 \times \frac{0.10\,\hat{\beta}_c}{\bar{y}_c}, \tag{A2}$$

ˆ其中， β_c 是入学率（选课学生人数除以分母）对儿童贫困率（比例）的加权斜率， $\bar{y}_c$ 是汇总后的全国比率，即该框架中选课学生总数除以分母总数； g_c 表示儿童贫困率每增加十个百分点，课程入学率相对于全国平均值的变化百分比。分母为课程所在年级的估算入学人数。对于高中课程，分母是9至12年级的估算入学人数（图1的分母）：每所提供任何9至12年级课程的学校，其民权数据收集（CRDC）总注册人数按该校K–12年级中高中年级所占比例折算。对于7至8年级的代数 I，分母则是7年级和8年级的入学人数，估算方法相同。9至10年级和11至12年级报告的两项代数 I 人数，均分别除以9至12年级的完整估算人数。这些分母反映相关年级的学校总体规模；它们无法识别哪些学生已满足课程的先修条件，因此该比率表示每名估算学生对应的选课人数，未衡量个人资格。文中还报告了人均变体以供比较，其分母为5—17 岁常住儿童人数（SAIPE）。区间为95%百分位区间，依据州聚类自助法进行，包含5,000个重抽样样本；每次重抽样均有放回地抽取州，并重新估算斜率和全国平均值，课程之间的差异以及课程排名也在各重抽样样本内计算。对每名常住儿童的估算高中入学人数采用同一估计量，这也解释了人均变体。稳健性变体重复上述计算，分别采用2015–16年和2021–22周期；采用中性框架，纳入所有报告了任一高中核心数学或科学课程入学人数的学区，并保留数据不完整的学区；以及加入州固定效应。州内斜率使用异方差稳健标准误，因为每个州仅提供一个聚类（图2）。

县级面板。图2的B面板重现了图中贝克萨尔县十五个学区的分布和梯度；若某学区半数以上的学校位于该县内，该学区即被视为属于该县。与A面板一样，每项比率均除以该课程的全国比率，因此两个面板中的1.00均表示全国比率。一个县的学区数量太少，无法按十分位划分：按儿童贫困程度依次排列学区，并将其合并，直至某一组至少包含该县估算高中在校生人数的十分之一；末尾不足一成的部分则并入最后一组。梯度是对十五个学区中比率关于儿童贫困程度进行回归所得的按入学人数加权的斜率，再除以该课程的全国平均值。其95%区间是通过自助法重抽样学区（即儿童贫困程度发生变化的层级）得到的百分位区间（5,000个重抽样样本）。十五个学区构成的聚类数量较少，因此这些区间只是近似值，且范围较宽。

地理聚合。 区域层面的背景变量通过哈希关系文件从较小的地理单元聚合到较大的地理单元，绝不反向聚合。机会地图集的普查区数值通过2010年人口普查的ZCTA–普查区关系 [49], ，按总体加权平均值聚合到邮政编码制表区；并通过美国国家教育统计中心教育人口与地理估算（EDGE）2017年地理

关系文件 [42]（TIGER 2017学区边界，代表2015–16学年，基于2010年人口普查区定义）按土地面积份额聚合到学区；每个聚合变量均保持生态层面的属性，并在模式中据此标记。任何一个文件都不会被用于将学生分配到某个普查区或学区（图1）。

排名持久性与一致性。 CRDC 周期之间的开课稳定性通过斯皮尔曼秩相关进行测量，即比较各学校在不同周期中某门课程的人均学生选课率。需要二元一致性统计量时使用 Cohen's κ，该统计量揭示了一个周期中退化的开课指标（附录 C.3）。

加权分布。 生源学区贫困状况分析使用加权直方图和加权中位数：对相同的学区层面贫困率分别按照送出的入学者人数和留在当地的儿童人数加权，因此两条曲线的差异仅来自加权方式（图 B2）。

A.4 机构描述性统计

流矩阵与环形布局。 专业变更流是基于所有专业并集构建的方形捐出方—接收方计数矩阵。总流量最小的十四个专业被归入一个标为"其他"的扇区；其中任意两个专业之间的流量将成为该扇区内部的流量。图中的带状图元以宽度表示学生人数，因此少于十名学生的流量不予绘制（共 610 条流量，合计涉及 2,325 名学生）。每个扇区的弧长与其在所绘流量中捐出和接收的学生人数之和成正比；扇区按该规模排序，因此没有任何带状图元、弧线或索引表行涉及少于十名学生。扇区以专业名称标注，因为一个专业可能对应多个教务处课程代码，而没有任何单一代码能够代表该专业。环形图整体刚性旋转，使最长的名称位于 135 度处；角度从 3 点钟方向逆时针测量，表D1的角度索引适用于弦图（图3）。

留一法偏差与层次聚类。对于学科 s 和初始专业 m（流程图中的来源专业），令 C_m 表示第一专业为m的专业变更学生集合。偏差矩阵项为

$$D_{s,m} \;=\; \frac{1}{|C_{s,m}|}\sum_{i \in C_{s,m}} \big(G_{i,s} - \bar{G}_{i,-s}\big), \tag{A3}$$

其中 $G_{i,\,s}$ 是学生 i在学科 s中的平均绩点（GPA），计算范围为计入教务处成绩平均值的课程行（退课和学分或无学分课程行不计入），$G_{i,\,-s}$ 是同一学生在所有其他学科中的平均GPA，而 $C_{s,\,m} \subseteq C_m$ 是在学科 s中有观测值的子集；少于15名学生的单元格留空。每个学科根据其观测单元格，以两种尺度进行汇总：学生加权平均偏差（每名专业变更学生的GPA分数）以及该平均偏差乘以观测到的专业变更学生人数（所有专业变更学生的总计）。行和列按照基于欧氏距离的Ward层次聚类排序，边缘处显示树状图（图4）。

分箱与 Hodrick–Prescott 平滑。 按教学班计算的平均绩点以每班人数相差 5 人为间隔进行分箱，在共同网格上插值，并使用 Hodrick–Prescott滤波器进行平滑；该滤波器是以下目标函数的最小化器

$$\sum_{t}(y_t - \tau_t)^2 \;+\; \lambda \sum_{t}\big(\tau_{t+1} - 2\tau_t + \tau_{t-1}\big)^2, \tag{A4}$$

使用 $\lambda = 10^4$ 求解为五对角线线性方程组。随着 $\lambda \to \infty$ 增大，滤波结果会趋近于一条直线，相当于线性回归，因此所选值决定了图中保留多少曲率（图5）。

入门课程结果与专业变更。 对于每个学科，将参考队列中第一学年修读过编号低于 2000 的课程且有首次修读记录的入学者，按这些课程的首次修读是否以 D、F、退课或无学分结果（DFW）告终分为两组，成绩代码与关联模型中的代码一致。比较两组中更换专业的比例（观察窗口内首次记录的专业与最后记录的专业不同）：计算比例差及其 Wald区间，并用对数（Woolf）法计算优势比及其区间。并列报告两个总体：入学后至少一年仍有入学记录的入学者，其定义为在入学学期之后一年或更晚的学期有注册记录，因此，暂时停学后返校的入学者也符合条件；以及所有有此类首次修读记录的入学者（表 B4）。

入门课程结果与持续就读。 针对各学科入学者的同一分组，比较其持续就读情况。结果指标与上述规则相反：如果从入学学期之后一年或更晚的学期起，直至观察窗口结束，记录中都没有课程，则视为后来未被观察到。课程结果发生在第一学年，因此早于这一结果。比较时剔除标记为协调招生计划的入学者，因为按照该计划的设计，他们会在第一学年后转学。还从两个方面进一步考察数学：按结果类型区分（已完成课程中获得 D 或 F 等，或仅有退课或无学分结果，因为退课可能是已经开始离校过程的一部分），以及入学者是否在另一门入门学科中也有 DFW结果（表 B3）。此外，还逐门课程统计 DFW结果：第一学年入门课程中每一次以 DFW 告终的首次修读均计为一个结果，因此入学者在同一学科中可能有多个此类结果。分别在所有入学者和后来未被观察到的入学者中，计算这些结果落在各学科的比例；并按入学者所持有的 DFW结果数量，计算后来未被观察到的入学者比例（图B6）。

队列限制优势比。 针对每个学科，在参考入学队列中，将暴露（在观察窗口内修读该学科）与结果（在同一观察窗口内注册高年级课程，不考虑所得成绩）制成 2×2 列联表，并进行 Fisher精确检验，使用对数（Woolf）法计算 95% 置信区间；在检验学科系列时，采用 $q = 0.001$ Benjamini–Hochberg 控制法。朴素对照则针对观察窗口内的所有学生计算同一列联表，仅作为附录 B.6 的设计比较报告（表B6和表B2）。

双样本检验。 当展示的统计量为平均绩点时，使用 Welch t t 检验（方差不齐）比较各面板，因为成绩分布大量集中在 4.0，秩检验在此情况下难以检出实际的均值差异；当沿用先前运营研究中的秩和惯例（每位捐赠者的平均绩点比率）时，则使用曼–惠特尼检验。两种情况下均采用 Benjamini–Hochberg 校正。

局部加权回归和质心。 全国完成率预期是对以本科教育为主的院校中，完成率关于佩尔助学金占比（接受联邦需求型佩尔助学金的本科生占比，这是一项标准的低收入指标）的局部加权散点平滑（比例为 0.4），并在院校的佩尔助学金占比处进行评估，从而得到佩尔助学金占比条件下的描述性残差；该残差

没有不确定性带，也没有协变量调整。每个面板由两个质心概括：按入学人数加权的院校质心（按当前本科生入学人数对位置均值加权，同时完成率坐标仍表示首次入学全日制学生队列结果）以及院校间的未加权均值；二者均在原始坐标中计算，并以显示尺度绘制（图 B5）。

A.5 联结入学者模型：设定

第 2.4 节说明了关联入学者模型的总体、测量指标和假设。本小节给出预测变量、结果和模型的完整设定。

预测变量。 家庭资源：学生是否申请助学金、申请者调整后总收入（AGI）在入学队列中的百分位排名，以及第一代大学生身份；所有秋季首次入学者中有 14.4% 未申请助学金，这些学生没有记录的收入，并单独构成一个类别。数据提取文件中每名学生有一个助学金申请标记和一个 AGI 值，但没有用于识别资助年度的字段，因此收入是一个无法核实时间的家庭资源标记；排名转换改变了其单位，但保留了记录中的时间信息。附录 B.4 在不使用资助变量的情况下重复报告主要估计结果。人口统计特征：性别和族裔。学区：入学前四个 SAIPE 数据版本的学区儿童贫困率平均值。学校：生源学校在高中毕业年级前后、距该年级一年以内的最近 CRDC 周期（2013–14、2015–16 或 2017–18）中，每名学生修读微积分、AP 数学、AP科学和双重注册课程的人数，以及总注册人数。存在一个例外。2013–14 周期仅将双重注册记录为指标，因此，对于分配到该周期的入学者，双重注册率采用 2015–16 年的比率，即最多晚于其高中毕业年级两年。高阶数学指数 是标准化微积分和大学先修数学比率的均值。这些学校变量描述的是入学者高中毕业年级前后生源学校记录的课程选修情况；它们不同于学校所开设的课程，也不同于某位入学者个人完成的课程作业。准备程度：入学数学和语言成绩，以及任何大学前数学学分的取得途径（无；仅双重注册；考试、AP 或大学先修课程考试计划 CLEP）。每项成绩均为院校记录的 SAT 和 ACT 中最高分科成绩；其中，ACT 和 2016 年以前的 SAT 均由院校转换为现行 SAT 量表；数据提取文件仅保留转换后的数值，没有考试日期或分数换算版本。我们在每个入学队列内，从每个数学成绩中减去队列均值，再除以该队列的标准差；语言成绩则以相同方式单独进行标准化，因此一个单位代表一个特定于队列的标准差（合并各队列时，数学成绩为 71 点）。其他连续预测变量均在关联样本上进行标准化。记录中没有高中平均成绩、班级排名或成绩单。

结果。入学准备程度通过两种方式衡量：标准化数学成绩，以及标准化数学成绩减去标准化语文成绩所得的差值，后者反映入学者在两项测试中的相对位置。第二种测量方式是两个分别标准化的分数之差，与教学班得分的原始差值不同。第一年表现以入学秋季学期及其后的春季和夏季学期中低年级课程的首次修读情况衡量：成绩（计入平均成绩的各行所对应的质量分数）和 DFW 指标（D 等、F 等、退课或无学分结果）。因此，表现取决于入学者所修读的课程。当至少有 1,000 名关联入学者在某一学科中有首次修读记录时，该学科即纳入分析；共有 25 个学科符合条件，其中 7 个在估算前被归类为定量学科（MAT、CHE、ECO、STA、PHY、CS、ACC）。成绩模型使用有成绩结果的首次修读记录，因此

成绩模型中的学生人数少于该学科达到纳入标准时的计数，其中一个学科的人数略低于 1,000。

模型。我们采用限制性最大似然拟合嵌套随机截距模型（学区，以及学区内的学校），将每项学业准备结果的方差分解为学区间差异、同一学区内学校间差异，以及同一学校入学者之间的差异。随后按顺序纳入各组预测变量，并报告每个方差成分的变化。我们将每位助学金申请者的收入排名拆分为同一生源学校入学者的平均排名，以及该学生相对于这一平均值的偏差。第一个系数比较来自平均收入排名不同的学校的学生；第二个系数比较来自同一生源学校的入学者，他们可能在不同年份就读。两者使用相同的收入排名尺度。这是协变量的组间与组内分解 [50]。两个系数之差描述的是情境关联，并不能确定就读某一特定学校的影响。每个学科模型都是采用课程与学期固定效应的最小二乘回归，因此，学生是与同一学期修读同一课程的其他学生进行比较；标准误按生源学校聚类，并对各学科采用 Benjamini–Hochberg 假发现率控制法，控制水平为 $q = 0.05$ [51]。各学科的系数通过随机效应进行合并 [52]。相同的学生和生源学校会纳入多个学科模型，因此，合并区间使用学科系数的学校聚类联合协方差：每一行对某个学科系数的影响函数按生源学校求和，这些和在不同学科间的交叉乘积则给出学科估计值之间的协方差。方案规定在独立性假设下进行合并；联合协方差是冻结后作出的修正（偏差 D2）。

样本外预测能力通过两项预测任务中的线性模型、岭回归和随机森林进行衡量。第一项任务中，十折交叉验证每次留出完整的生源学校，预测目标是学生成绩相对于同一课程、同一学期平均成绩的差值。该平均值仅根据训练折计算，因此留出结果不会进入目标值或预测；对于课程和学期在训练折中均无对应记录的留出行，则采用该课程在训练折中的平均成绩（占总行数的 0.03 %）。这项任务描述的是，在已知某课程开课学期的平均结果时，入学特征对学生在该课程中的排序效果。它属于回顾性分析，因为课程和学期的平均值要到学期结束后才能获得。第二项任务中，模型使用 2014 年秋季和 2015 年秋季的入学者数据进行拟合，并预测 2016 年秋季至 2018 年秋季入学者的成绩；基线为较早入学届中同一课程的平均成绩（对于较早入学届未修读的课程，占行数的 1.9%，则采用该学科的平均成绩）。这项任务的所有结果基线均根据训练队列估算。该任务评估跨队列预测能力。数据摘录无法确定每个记录的预测变量何时可用：分数字段没有考试日期，助学金字段也没有资助年度。在两项任务中，预测能力均以相对于基线的平方误差比例降幅衡量；某一组预测变量的贡献，则是其按顺序纳入时增加的预测能力，以及将其剔除时损失的预测能力。方案在划分训练折之前，使用了所有行计算课程与学期的平均值；仅使用训练数据构建均值是冻结之后作出的修正（偏差 D2），附录 B.4 同时报告了两种结果。通过对生源学校进行重抽样来评估各组预测变量贡献的变异性，并对每个学科应用相同的抽样结果，同时保持拟合预测不变。所得区间描述的是在给定这些预测结果的情况下评估样本中的变异性；其中不包括重新拟合和调优模型所产生的变异性。

A.6 教学班规模模型：入学规则、推断和检查

第 2.4 节说明了班级规模模型的比较和假设。本小节介绍纳入规则、推断和检查。

纳入规则。 当某一学科至少有 200 名学生拥有符合条件的记录，且移除课程×学期效应和教师效应后，班级规模仍有变化（对数的标准差至少为 0.20）时，该学科即纳入分析；25 个学科中有 20 个符合条件，其余 5 个学科的教学班规模固定，或样本稀少。

推断。 在每门课程和每个学期内打乱成绩后对分析流程进行试运行，结果显示，在这种设计下，使用按授课教师聚类的标准误进行的 t 检验过于频繁地拒绝原假设（对文件中校准运行涵盖的 5 个学科，每个学科打乱 200 次，在名义显著性水平为 5% 时，拒绝率为 5.5% 至 19.0%）。因此，每个班级规模项的 p 值都来自对授课教师进行的野生聚类自助法 t 统计量检验：检验时施加原假设，并进行 999 次抽样；这是 Cameron 等人 [39] 研究的一类方法。在相同的打乱设置下，该方法的拒绝率为 2.0% 至 6.5%。某项的 95% 区间是对称的：估计值加减自助法 t 统计量绝对值的第 95 百分位数与按授课教师聚类的标准误的乘积。对于正文重点讨论的量，我们还报告了在一系列原假设值上反演自助法检验所得的区间，二者十分接近。在给定入学分数下的班级规模斜率，以及入学分数分布三分之一处的情景，都是规模系数与交互项的线性对比。每个对比都作为重新参数化模型中的一个系数，用相同的自助法进行检验，因此其区间包含这两个项之间的协方差。方案为这些对比指定了聚类 t 区间；作为冻结方案后所作的修正，这些区间由自助法区间取代。在打乱的成绩下，这些对比的自助法检验在名义显著性水平为 5% 时，拒绝率为 3.5% 至 7.5%（打乱 200 次）；相比之下，聚类 t 检验的拒绝率为 4.5% 至 9.5%。学生固定效应模型对其 758 名授课教师仍采用聚类区间。对于各学科，采用 Benjamini–Hochberg 方法控制 $q = 0.05$，并应用于自助法 p 值。合并系数和总计所用的区间纳入了学科估计值之间因授课教师和学生重叠而产生的依赖关系，该依赖关系通过分三步构建的协方差矩阵纳入。每一行对其所属学科系数的影响函数，分别在授课教师和学生内求和，由此得到学科之间的双向聚类协方差 [53]。随后将该矩阵转换为相关矩阵，并将负特征值设为零，从而投影到最近的半正定矩阵。再以每个学科的自助法标准误（其区间半宽除以 1.96）缩放相关系数。因此，最终结果是一种混合方法：跨学科相关性来自双向聚类，边际不确定性来自野生聚类自助法。估计的跨学科相关性接近于零，因此总计几乎没有变化。关联入学者模型的合并区间采用另一种构造方法，即上文所述的按学校聚类的联合协方差。

冻结前观察到的检查与指标。由于较小教学班中的较高成绩可能反映评分较宽松，因此，在给定首次成绩的情况下，下一门数学课程的成绩可作为一项有限的敏感性检验：它只使用继续修读另一门数学课程的入学者，并以首次成绩为条件。两个指标经打乱后仍然存在，且在冻结前已被观察到：跨课程系数，以及入学者修读后续数学课程的比率；后者作为描述性指标报告。在查询内部，教师标识符会被替换为代理标识符，且没有任何结果按教学班或教师编制索引。

B 补充分析结果

本附录报告支持第 3 节结果的分析。每项情境分析均遵循独立方案：每幅图使用一个数据源（或一组经过审计的汇总数据源），图注中注明数据源总体，分析单位保持在数据源自身的层级，并且绘图数据集中不包含任何行级

学生记录。两幅背景图（图 B1 和图 B2）根据入学人数对学校层级或学区层级汇总数据进行加权，其估计目标在此意义上仍以入学者为条件。在背景图中，共同地理区域绝不会建立个人层级关联。学生记录会根据第 2.4 节的规则进入附录 B.4 的关联入学者模型和班级规模模型。本附录中的任何内容都不支持因果解读。

B.1 学区课程梯度：稳健性与人均指标

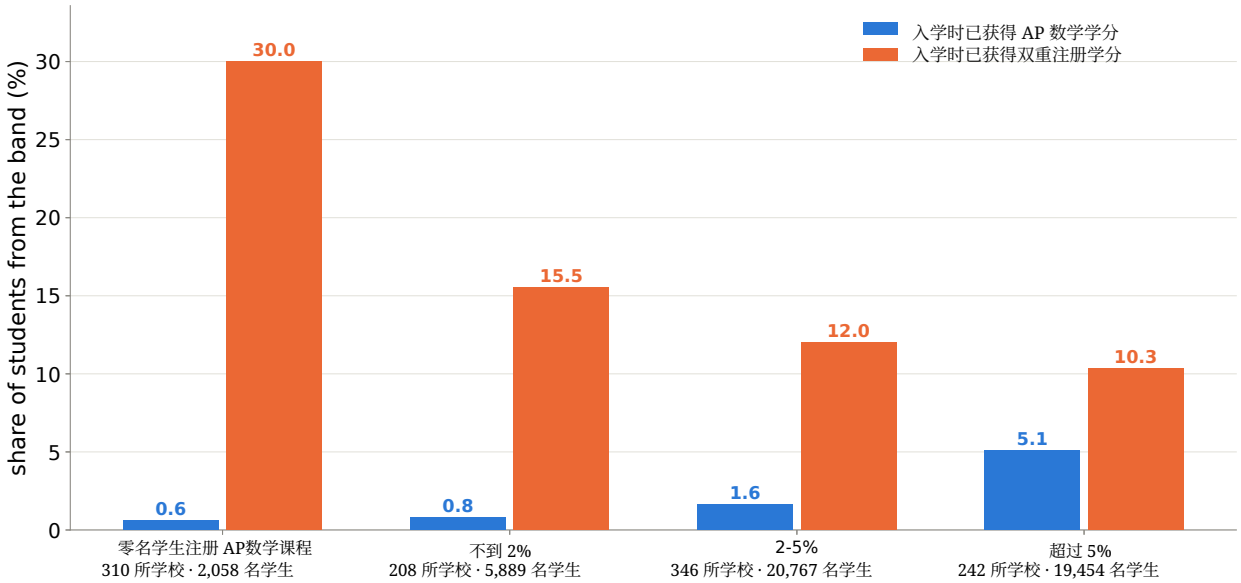
本小节为第 3.2 节提供支持。

六项检查支持这一对比。在所纳入的总体周期中，三门选修人数众多的核心课程的梯度均保持在零点上下四个百分点以内（2015–16年、2017–18年和2021–22年分别介于 -1.4% 与 $+3.4\%$ 之间），而微积分梯度分别为 30.7%、36.9% 和 45.3%。保留总数不完整的学区（2017–18年有五个）后，这些梯度的变动均不超过半个百分点。采用中性框架，纳入所有报告了任意一门高中核心数学或科学课程的学区，微积分比例为 37.1%，选修人数众多的核心课程则介于 -0.6% 与 $+1.4\%$ 之间。加入州固定效应后，微积分梯度为 37.0%，选修人数众多的核心课程仍在零点上下三个百分点以内。分母检验仅限于年级范围恰为 9–12、无需按比例分摊入学人数的 13,250 所学校，结果显示微积分比例为 37.5%（区间为 29.1 至 47.0 ），选修人数众多的核心课程则在零点上下两个百分点以内。此外，在得克萨斯州（ 860 个学区），AP 数学（35.7%）、微积分（24.7%）以及 7至8年级的代数 I（21.5%）均大幅下降。得克萨斯州还提供了一项政策检验。第5号众议院法案规定，除要求修读代数 II 的专业方向外，代数 II 均为选修课；2017–18年是所有四个高中年级首次纳入基础高中课程计划的周期 [31], ，且在得克萨斯州，代数 II 的选课人数与高中生人数之比没有贫困梯度（$+0.4\%$，异方差稳健 $z = 0.30$）。得克萨斯州仅构成一个聚类，因此这些是州内描述性斜率。

人均指标呈现出另一番情况，原因与课程无关。以 5—17 岁常住儿童人数为分母，高中在校生人数估计值本身每当儿童贫困率上升 10 个百分点就下降 8.7%（区间为 6.1 至 11.7）；这一降幅也体现在每门课程上：按每名常住儿童计算，广泛开设的核心课程似乎以与高中在校生人数相同的幅度下降（7.7% 至 9.4%），微积分和 AP 数学与科学课程的降幅则达到 42% 至 43%。以各课程相应年级的学生人数为分母进行衡量，就能排除入学人数的地理差异，单独考察课程本身。

B.2 输送学校课程供给与入学时的学分

图B1展示了生源高中提供 AP 数学课程的情况，统计对象为48,168 名在读学生；这些学生的高中均可对应到2017–18 年民权数据收集中的一所学校（共1,106所）。从左向右看，来自未记录有 AP 数学课程选修的学校的学生，入学时获得 AP 学分的比例为0.6%，获得双重注册学分的比例为30.0%；来自 AP 数学课程选修人数占比超过5%的学校的学生，这两个比例分别为5.1%和10.3%。随着课程供给分档变化，这两项入学占比呈相反方向变化；零课程供给学校的规模较小（在校人数中位数为266人，而最高档为1,598人）。该图根据学校分档汇总数据构建入学概况，未按学校规模或学生特征进行调整；附录 B.4 中的关联模型则在调整后分析入学准备程度与生源学校背景的关系。此处的课程供给数据是2017–18年快照，用于描述高中就读年份大多早于该时点的学生。从2015–16年至2017–18年各周期，生源学校在人均比率方面的排名并非完全稳定（AP科学的秩相关为0.78，AP 数学为0.76，微积分和双重注册均为0.62）；因此，这一快照反映的是较晚时期的学校背景，只有在课程供给保持稳定的情况下，才能近似代表各入学者本人高中时期的课程供给。



图B1：按生源学校提供 AP 数学课程的情况划分的入学学分分布，统计对象为 2014年秋季至2019年秋季期间在读且高中信息已确认的学生。根据各生源学校自身在读学生中修读 AP 数学的占比（CRDC 2017–18）对学校分组；柱形图显示各组中入学时获得 AP 学分（蓝色）和双重注册学分（橙色）的学生占比；每组下方标注学校和学生人数（总计 1,106 所学校、48,168 名学生）。各组中 AP 学分的入学占比从 0.6% 上升至 5.1%，而双重注册学分的入学占比从 30.0% 降至 10.3%：两种入学学分占比随分组变化呈相反趋势。课程供给情况定义为该周期内记录的零<击数 id='1'>入学人数，并在学校层面进行描述；本图展示的是生源学校背景，绝不将其视为学生接触史。每个单元格至少包含十名学生。

B.3 生源学区的儿童贫困

图 B2描述了向该院校输送学生的学区中儿童贫困的分布情况，统计对象为53,367名在读学生；这些学生的高中均可对应到得克萨斯州公立学区（共682个学区）。按各学区输送的学生人数加权，生源学区儿童贫困率的中位数为16.3%；按各学区儿童人数加权，得克萨斯州的中位数为21.9%。因此，该院校的入学群体比全州按儿童人口加权的中心低约5.6点；其中15–17.5%这一档涵盖了36%的学生，因为两个主要生源学区——北区独立学区（9,573名学生，按2010–2018年平均值计算的儿童贫困率为16.3%）和 North East ISD（6,371名学生，儿童贫困率为15.8%）——均属于这一档（排名靠前的生源学区列于该图的配套文件中）。在样本中，无法匹配到学区的学生比已匹配学生更常获得佩尔助学金；这一差异处于不同生源学校之间的变异范围内。无法观察这些学生所在学区的贫困程度，因此无法确定已匹配学生的分布与全体学生分布之间差异的方向（附录 C.5）。这两种分布均为学区层面的 SAIPE 比率汇总；分析中未纳入学生个人收入。

B.4 从生源背景到准备程度和大一表现

背景分析描述了学区和学校。本小节转而探讨学生。在2014年秋季至2019年秋季的23,489名关联入学者中，本节估计家庭、学校和

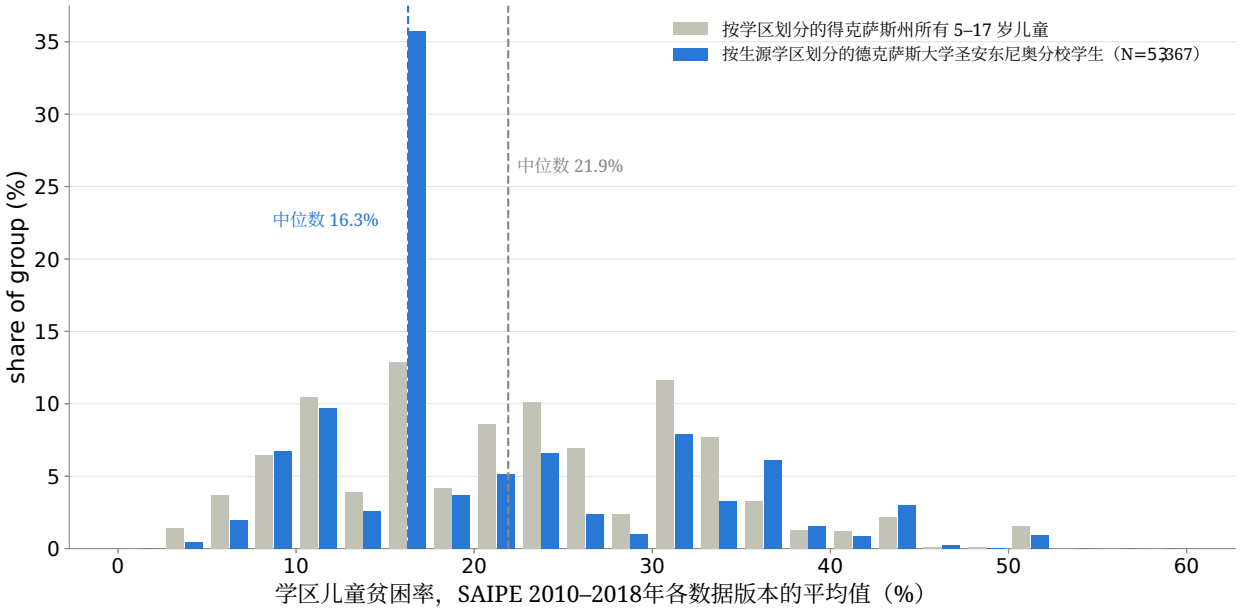


图 B2：生源学区的儿童贫困情况。灰色：得克萨斯州所有 5–17 岁儿童的学区儿童贫困分布（SAIPE 2010–2018年各数据版本的平均贫困率），按各学区所辖儿童人数加权。蓝色：2014年秋季至2019年秋季期间在读的 53,367 名学生的相同分布；这些学生的生源高中可对应到得克萨斯州公立学区（共 682 个学区），按各学区输送的学生人数加权。权重为汇总计数；本图未使用任何学生层级记录。虚线表示加权中位数。本图涵盖了拥有得克萨斯州高中名称的 60,045 名学生中的 88.9%。其他学生的学区贫困程度无法观测，因此蓝色分布仅描述已匹配的学生。

学区特征与入学数学准备程度之间的关联，以及准备程度与25个学科中各学科第一年表现之间的关联。入学数学成绩的一个标准差为71点；所有估计均以入读该院校为条件。该方案在首次针对11个入学队列运行后，才采用对这些队列的限制（第 2.4 节）；本小节最后几段将说明这一限制对研究设计的影响。

准备程度存在差异之处。 控制入学队列和 CAP 状态后，随机截距模型将数学成绩剩余方差的 7.9 % 归因于学区之间的差异，14.3% 归因于同一学区内各学校之间的差异，合计为 22.2%。其余 77.8% 是同一所学校入学者之间的变异（图 B3）。这一划分描述了分数在哪些层面存在差异，但无法确定差异的成因。将家庭资源、人口统计特征和学区儿童贫困纳入模型后，学区间差异部分减少了 95%，学校间差异部分减少了 35%；加入学校模块后，学校间差异部分的降幅达到 63%。在完整模型中，学校高阶数学指数每增加一个标准差，对应 $+9.7$ 个分数点；学区儿童贫困每增加一个标准差，对应 -10.4 个分数点；调整后总收入排名每增加一个标准差，对应 $+5.9$ 个分数点；第一代大学生入学者的成绩比非第一代大学生入学者低 10.6 分。由于数据摘录未记录助学金变量的时间信息，因此不纳入这些变量时，两个背景系数分别为 $+9.8$ 和 -11.5 个分数点。

学校间与学校内的收入差异。 收入与数学成绩的关联在生源学校之间明显大于学校内部：学校间收入排名每增加一个标准差，成绩相差 0.436 个标准差；同一所学校的入学者之间，每增加一个标准差，成绩相差 0.048 个标准差，两者相差 0.39（$z = 10.2$）。这一差异与学校间的社会经济分层相符。但这一比较无法区分学校、家庭、社区或进入大学的选拔所产生的影响。

相较于语言成绩的数学成绩。对于标准化数学成绩减去标准化语言成绩这一差值，97.5%的变异来自同一所学校的入学者之间，2.5%来自学校和学区之间。根据预先设定的判定规则，H1 得到支持：在控制家庭资源后，学校的高阶数学指数与这一指标呈正相关，但系数较小：$+0.027$ 个标准差/指数每增加一个标准差（区间为0.008至0.046）；若数学成绩的一个标准差按71点计算，则相当于1.9点。在11项敏感性分析中，估计值均为正，其中两项的区间包含零（仅纳入2015年秋季及之前的入学者，以及仅纳入CAP入学者）；不纳入助学金变量时，结果不变（$+0.027$，0.008至0.046）。在同一所学校的入学者中，收入排名与这一指标无关联（-0.01）。系数最大的是国际学生（相较于西班牙裔学生为$+0.72$个标准差，标准误为0.17）、亚裔学生（$+0.39$）和男性（相较于女性为$+0.36$）。排除生源学校后，预测变量可以解释数学成绩29%的变异，以及两项成绩之差6.5%的变异。在这些入学者中，学区贫困程度和学校背景与两项成绩的水平相关，但与二者之间的差值仅有较弱关联。

准备程度与一年级成绩。 表 B1 和图B4报告了学科模型，这些模型比较了在同一学期修读同一课程的学生。图中的每一行代表一个学科。前七列来自包含学生特征的模型，最后两列来自加入学区和学校学期项的同一模型。每个单元格表示与该列特征相对应的绩点差异，同时保持模型中其他特征相同。例如，左上角的单元格表示，在数学课程中，若一名入学者的入学数学成绩比同一课程、同一学期的另一名学生高一个标准差（71点），则其平均成绩高出0.38点。H2得到支持：在控制语言成绩后，在错误发现控制下，数学成绩在18个非定量学科中的17个以及全部7个定量学科中均具有正系数。在定量学科之外，数学成绩每增加一个标准差，与绩点提高0.140相关（区间为0.109至0.171）；相比之下，语言成绩对应的绩点提高为0.112，二者差异的区间包含零（$+0.027$、-0.015 至 $+0.068$）。在定量学科之内，数学系数为0.263个绩点（0.197至0.328），语言系数为0.065。这些合并值赋予每个学科相近的权重。按首次尝试加权后，定量学科之外的数学系数为0.125个绩点，定量学科之内为0.316；在后者中，数学既是规模最大的学科，也是系数最大的学科；数学系数通过检验的24个学科占首次尝试的98%。在数学课程本身，数学系数为0.38个绩点，DFW概率为 -10.6百分点；语言系数略为负值（-0.03）。在四个学科中，语言系数高于数学系数，其中两个学科的差异通过了错误发现控制，即政治学和人类学。基于所有入学者估计的系数与关联样本相符；在每个学科中，逻辑回归模型与线性DFW模型的符号一致；在不包含资助变量的情况下，定量学科之外的合并数学系数为0.145，其中25个学科中的24个

如前所述通过。

相同的学生和生源学校会参与多个学科模型。上述区间使用了学科系数的联合协方差（第 2.4 节）；两个学科的数学系数之间的估计相关性平均为 0.08，最高达到 0.47。我们还对定量组之外各学科的数学系数与语言系数之差取平均，并按各学科的首次尝试次数加权。纳入协方差后，该平均值的区间包含零（-0.008 至 $+0.043$）。在方案所假定的独立性下，相同区间不包含零（0.001 至 0.035），而定量学科之外数学系数的合并区间为 0.111 至 0.169。因此，这 24 个正向学科结果是基于重叠学生的相关观测值，数据并不能支持在定量学科之外将数学准备程度排在语言准备程度之上。

学区和学校变量的预测贡献。 在第一个预测任务中，生源学校被留出，目标是学生相对于同一课程和学期平均成绩的成绩；该平均值根据训练折计算。完整的预测变量集合使各学科的平方误差中位数降低 8.6%（范围为 2% 至 19%；按首次尝试加权时为 10.3%）。准备模块（分数和学分路径）对平方误差比例下降的贡献中位数为 0.056（区间为 0.044 至 0.065；该区间通过对生源学校进行重抽样获得，并保持拟合预测不变）；家庭资源的贡献为 0.003，人口统计特征的贡献为 0.008。学区贫困程度的贡献中位数为 0.0004（区间为 -0.0008 至 $+0.0016$），在各学科之间介于 -0.004 和$+0.003$ 之间，并在 25 个学科中的 17 个学科中为正。学校模块的贡献中位数为 -0.0015（区间为 -0.0033 至 0.0000），在各学科之间介于 -0.013 和 $+0.003$ 之间，并在 7 个学科中为正。方案规定的评估使用所有行计算课程—学期平均值，所得数值几乎相同（完整集合为 8.5%，准备程度为 0.057，学区贫困程度为 0.0004，学校模块为 -0.0022 ）。在第二个任务中，根据 2014年秋季和 2015年秋季入学者拟合的模型，使用每门课程较早时期的平均成绩来预测后三个队列的成绩。岭回归使平方误差中位数降低 8.35%（范围为 1% 至 17%），而未加惩罚的线性模型由于在较小学科中过拟合，降低 6.9%；学区贫困程度对线性模型的贡献中位数为 0.001，学校模块为 -0.004。对于 DFW结果，预测概率在第一个任务中的曲线下面积中位数为 0.67，在第二个任务中为 0.63，而仅使用基准率时分别为 0.58 和 0.52。

因此，在入学成绩、学分路径和其他学生特征的基础上，加入学区贫困程度和学校课程注册变量后，预测准确率中位数几乎没有变化，而且变化方向因学科而异。这一结果仅适用于现有的预测变量和这两种评估设计。它并不能证明学校背景与表现无关：在若干学科中，学区贫困程度的调整后系数为负，且区间不包含零（图B4）；系数与预测准确率的增量衡量的是不同的方面。随机森林在第一项任务中的表现并未比线性成绩模型高出超过0.009，在第二项任务中的表现也未比岭回归高出超过0.008。我们未发现这种替代模型带来更大的提升；这一比较并不能排除随机森林同样未能捕捉到的交互作用或曲率。

其他预测变量。 在相同课程和学期内，CAP入学者在25个学科中的18个学科的成绩比其他入学者高0.1至0.6点。考试学分（大学先修课程或CLEP）与17个学科中较高的成绩相关。双重注册学分与数学以外学科的较高成绩相关，而与数学课程的较低成绩相关（-0.17）；在数学课程中，考试学分系数为正（$+0.10$），但其区间包含零。第一代大学生入学者的成绩低0.07至

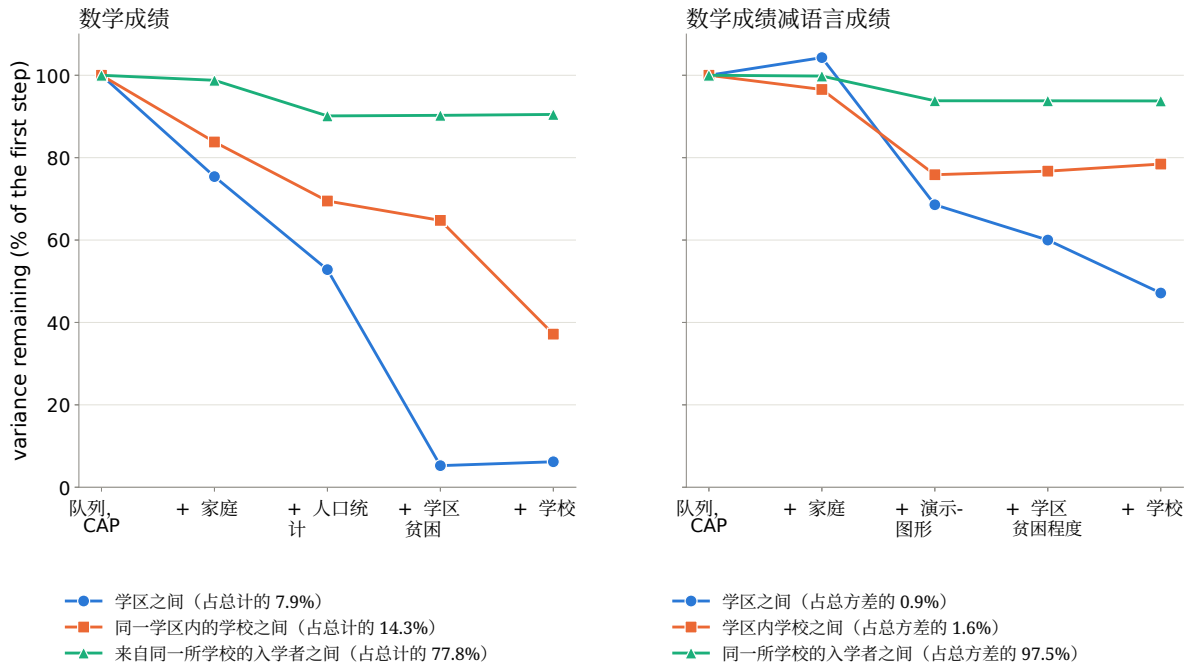
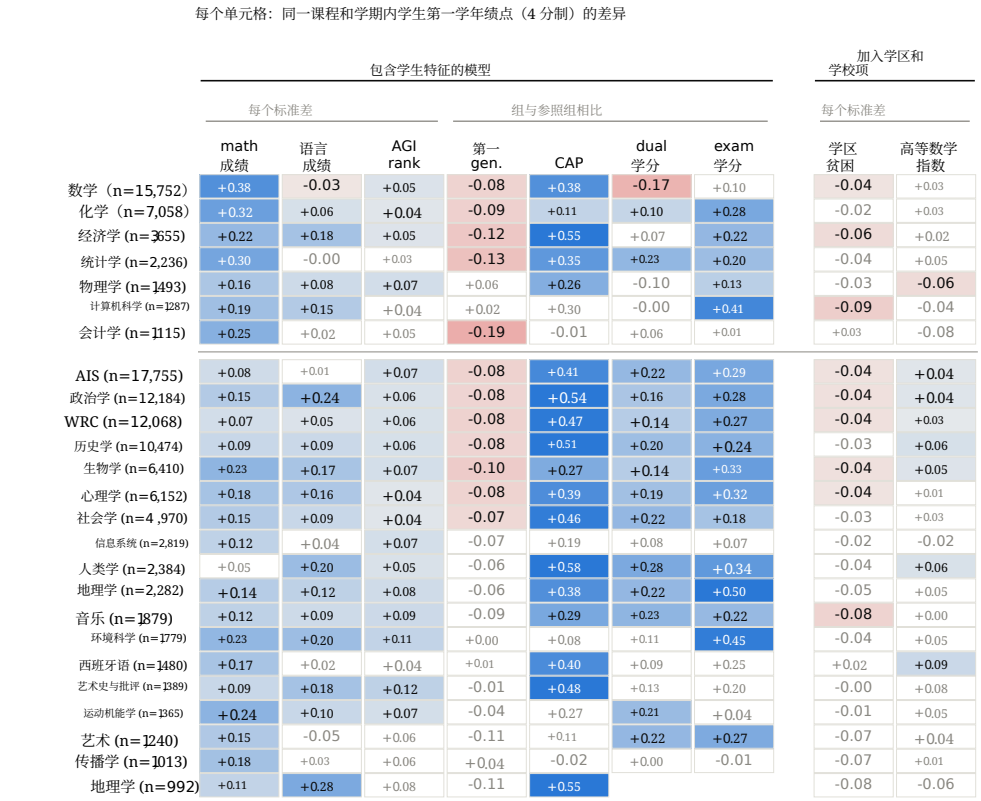


图 B3：2014年秋季至2019年秋季的 23,478 名关联秋季首次入学者（来自 529 个学区的 882 所生源学校）中，入学准备程度的差异分布在哪里，以及哪些预测变量组可以解释这些差异。每条线追踪嵌套随机截距模型中的一个方差成分，各组按以下顺序纳入模型：家庭资源（助学金申请、调整后总收入排名、第一代大学生身份）、人口统计特征、学区儿童贫困情况，以及学校模块（高阶数学指数、AP科学和双重注册率、入学人数）。左图：数学成绩，按入学队列进行标准化。右图：数学成绩减去语言成绩。图例中的百分比表示第一步中各成分占总方差的比例。所有数值均以进入该院校为条件。

0.19点，涉及的12个学科的区间不包含零（未对多重性进行调整）。

队列范围的限制。第 2.4 节说明了为何将入学者限定为2014年秋季及之后入学者：在方案先对11个队列进行首次运行并获知结果之后，才采用了这一限制。在研究设计中，这一限制使符合至少1,000名学生这一最低人数要求的学科从11个队列时的31个减少到25个；同时，每位入学者的学校概况与其高中毕业年级之间的间隔缩短至一年以内，而较早入学届中最长可达五年。有一个变量例外：对于被分配到2013–14周期的8,180名关联入学者，双重注册率采用的是2015–16年的数据，距其高中毕业年级最远可达两年。

范围。范围上的局限与第4.3节所述相同：入学选择，以及吸收一般学业准备程度的入学成绩。学科模型衡量已修课程条件下的首次修读成绩和DFW结果；入学者开始修读的课程本身也与其准备程度相关，而高阶数学指数对应于更高阶的入门数学课程（每个标准差对应五级量表上的+0.07个等级）。这些分析探讨学校背景与准备程度之间的关联，以及准备程度与成绩之间的关联；但并未估计贫困与成绩之间的关联有多少是由学校课程供给或准备程度造成或中介的。



图B4：2014年秋季至2018年秋季期间，在至少1,000名关联入学者中有首次修读记录的25个学科里，入学特征与第一学年成绩之间的关联。每个学科都有两个采用课程与学期固定效应的回归，因此比较的是修读同一课程、处于同一学期的学生：前七列来自包含学生特征的模型，最后两列则来自在同一模型中加入学区和学校项后的结果。每个单元格表示与该列特征相关的绩点差异（4 分制），并控制该模型中的其他特征；图中为左上角单元格提供了计算示例（0.38个绩点约相当于三分之一个字母等级成绩）。标注“每标准差”的列表示与该特征每增加一个标准差相关的差异。标注“群体与参照组比较”的列分别比较第一代大学生入学者与非第一代大学生入学者、CAP入学者与其他入学者，以及持有各类大学前数学学分的入学者与未持有任何此类学分的入学者。阴影单元格（蓝色为正、红色为负）的区间不包含零：分数列在 $q = 0.05$ 个学科范围内进行错误发现控制，其他列则未作调整$p < 0.05$。7个定量学科位于分界线之上。计数 n 表示有评分的首次修读记录的关联入学者人数，可能低于纳入该学科所需的1,000人（最少为992人）。标准误按生源学校聚类；每个单元格表示以入学为条件的关联。学科代码：MAT，数学；CHE，化学；ECO，经济学；STA，统计学；PHY，物理学；CS，计算机科学；ACC，会计；AIS，一年级研讨课；POL，政治学；WRC，写作；HIS，历史学；BIO，生物学；PSY，心理学；SOC，社会学；IS，信息系统；ANT，人类学；GEO，地质学；MUS，音乐；ES，环境科学；SPN，西班牙语；AHC，艺术史与批评；KIN，运动机能学；ART，艺术；COM，传播学；GRG，地理学。

表 B1：一年级学科模型中两个入学分数的系数，按各分数的一个标准差计，附 95% 区间。对于每种结果和学科类别，第一行是预先指定的跨学科随机效应合并，使每个学科都具有相近的权重；第二行则按各学科首次尝试的次数加权，对学科系数取平均，是以所服务入学者为基准的描述性汇总。成绩采用 4 分制；DFW 是获得 D、F、退课或无学分结果的概率，以百分点表示。同一批学生和生源学校会纳入多个学科模型，因此区间采用学科系数的按学校聚类联合协方差；这一修正是在方案冻结后作出的（方案原先将各学科视为相互独立）。学科按估计前的分类分为定量学科（MAT、CHE、ECO、STA、PHY、CS、ACC）和其他学科。总体：2014年秋季至2018年秋季入学的关联入学者；这一限制是在预先指定的十一届队列分析后采用的。

结果，学科类别（学科）	数学	语言	数学减语言
绩点，其他（18）	+0.140 (+0.109, +0.171)	+0.112 (+0.069, +0.156)	+0.027 (−0.015, +0.068)
按首次尝试加权	+0.125 (+0.108, +0.142)	+0.107 (+0.093, +0.122)	+0.018 (−0.008, +0.043)
定量绩点（7）	+0.263 (+0.197, +0.328)	+0.065 (+0.000, +0.129)	+0.197 (+0.073, +0.320)
按首次尝试加权	+0.316 (+0.291, +0.340)	+0.032 (+0.012, +0.053)	+0.283 (+0.244, +0.323)
DFW、点数、其他（18）	−2.2 (−3.0, −1.5)	−1.6 (−2.9, −0.3)	−0.7 (−1.8, +0.5)
按首次尝试加权	−2.0 (−2.5, −1.6)	−1.5 (−1.9, −1.0)	−0.6 (−1.3, +0.2)
DFW、点数、定量（7）	−5.1 (−7.8, −2.4)	−0.9 (−2.7, +0.9)	−4.2 (−8.3, +0.0)
按首次尝试加权	−7.7 (−8.4, −6.9)	+0.2 (−0.3, +0.8)	−7.9 (−9.0, −6.8)

B.5 该机构的全国排名

截至最近一个公布完成结果的数据年份（2024–25），德克萨斯大学圣安东尼奥分校报告本科生人数为 30,580，首次入学全日制学生在标准修业时间 150% 内完成学业的比例为 0.526，佩尔助学金占比为 0.441；根据 Scorecard 中以本科教育为主的院校数据，对完成率与佩尔助学金占比进行局部加权拟合后，该佩尔助学金占比所对应的条件预期值为 0.462，描述性残差为 $+0.064$ [54]。在中断前的数据版本（数据年份 2019–20）中，相同坐标对应的本科生人数为 27,573，完成率为 0.423，低于预期值 0.478，残差为 -0.055。需注意范围：德克萨斯大学圣安东尼奥分校与 UT Health San Antonio 自 2025年9月1日 起合并，因此 2024–25 这一行描述的是合并前的旧报告单位，而非合并后的院校 [8]。该残差衡量相对于全国预期值的排名；这两个数据版本之间变动的归因问题留待后续比较研究探讨。

B.6 比较组定义：设计比较

混合不同入学路径的比较衡量的是这些入学路径，而不是课程；必修一年级研讨课的结果证明了这一点。院校的存储过程会针对任何学科，在观察窗口内的所有学生中，计算修读该学科与未记录修读编号高于 4000 的课程（高年级课程）之间的优势比。对于必修研讨课 AIS，这种朴素设计得出的不利优势比为 1.81（表 B2）：修读必修研讨课的学生似乎表现更差。这个数值是比较组造成的假象。几乎所有首次入学新生都会修读 AIS（占参考队列的 92.6%），因此，朴素设计中的“未暴露”组几乎完全由转学生和晚入学者构成：前者携带在其他院校获得的学分，很快达到高年级水平；后者的观察窗口则受到删失。将两组都限制为具有固定观察窗口的参考入学队列后，关联转为明显有利（0.23），因为在该队列中，跳过必修研讨课的学生正是那些立即离校的学生；剩余的少量未暴露学生本身高度

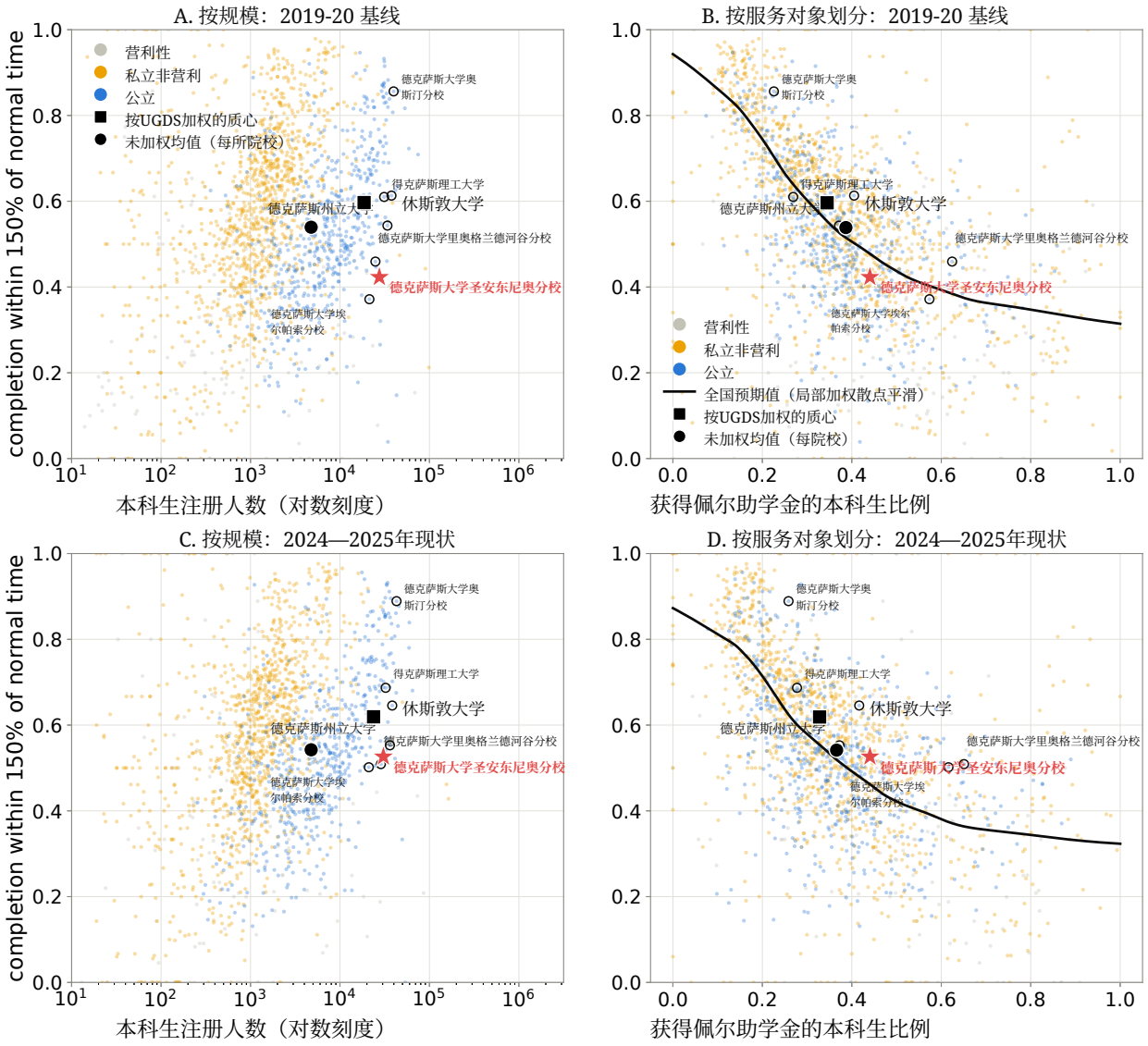


图 B5：在 College Scorecard 中以学士学位为主的第四章院校里，德克萨斯大学圣安东尼奥分校在两个明确的数据版本中的全国排名。上排（A、B）：2019–20 数据年份（1,869 所院校），即中断前的基线版本，其毕业队列在 2019年秋季之前结业。下排（C、D）：2024–25 数据年份（1,814 所院校），即 2026年6月发布版本中最新的毕业数据；本排报告的是合并前沿用的报告单位 [8]。左列：首次入学的全日制学生在标准修业时间的 150% 以内完成学业的比例与本科生注册人数（对数刻度）的关系。右列：完成率与佩尔助学金获得者占比的关系，并显示全国局部加权散点平滑预期值；德克萨斯大学圣安东尼奥分校与曲线之间的差距是佩尔助学金占比条件下的描述性残差。公立院校以蓝色表示，私立非营利院校以黄色表示，营利性院校以灰色表示；得克萨斯州公立同类院校以圆圈标出。黑色方块表示按入学人数加权的院校质心（按当前本科生在校人数对院校加权；完成率坐标仍为首次入学全日制学生队列结果）；黑色圆圈表示原始坐标中的院校未加权均值，每所院校各占一票，不论规模大小，此处绘于对数轴上。

表 B2：在两种设计下，核心学科暴露与观察窗口内未选修编号高于 4000 的课程之间的关联。朴素设计（第 2–4 列；在整个 2014年秋季至2019年秋季观察窗口内运行存储过程）将每门学科的修课者与窗口内所有其他学生进行比较，该群体混合了持有在其他地方获得的学分的转学生和右删失的晚入学者。队列设计（第 5–7 列）则将两组都限定为在 2014年秋季至2015年秋季入学的首次入学新生，并采用第 2.2节规定的固定观察窗口。优势比大于 1 表示修课者未选修此类课程的优势较高； p 来自 Fisher精确检验；95% 置信区间采用对数（Woolf）法计算。由 cohort tables.py 生成。

	朴素设计（观察窗口内的所有学生）			限定队列（入学队列）		
学科	OR	95% 置信区间	p	OR	95% 置信区间	p
MAT	1.47	[1.43, 1.51]	$< 10^{-4}$	1.46	[1.27, 1.67]	$< 10^{-4}$
WRC	1.80	[1.74, 1.85]	$< 10^{-4}$	1.39	[1.28, 1.52]	$< 10^{-4}$
AIS	1.81	[1.76, 1.87]	$< 10^{-4}$	0.23	[0.19, 0.28]	$< 10^{-4}$

因此，即使是队列限制的 AIS 估计，描述的也不仅是研讨课的影响，还包括这种选拔。数学和写作入门课程在同样的限制下表现不同：MAT 基本不变（朴素估计为 1.47，队列估计为 1.46），而 WRC 的不利关联有所减弱（从 1.80 降至 1.39 ）。这一限制消除了入学路径伪影，但并未对准备程度、专业、修读学分负荷或任何其他协变量进行调整；此外，暴露和结果是在同一观察窗口内测量的，因此这些仍是同期边际关联。这个例子说明了本文院校分析所遵循的规则：每项关联都基于明确界定的入学人群，并注明观察窗口（表 C1）；凡比较组混合了不同入学路径的统计量，均视为对这些入学路径的测量。

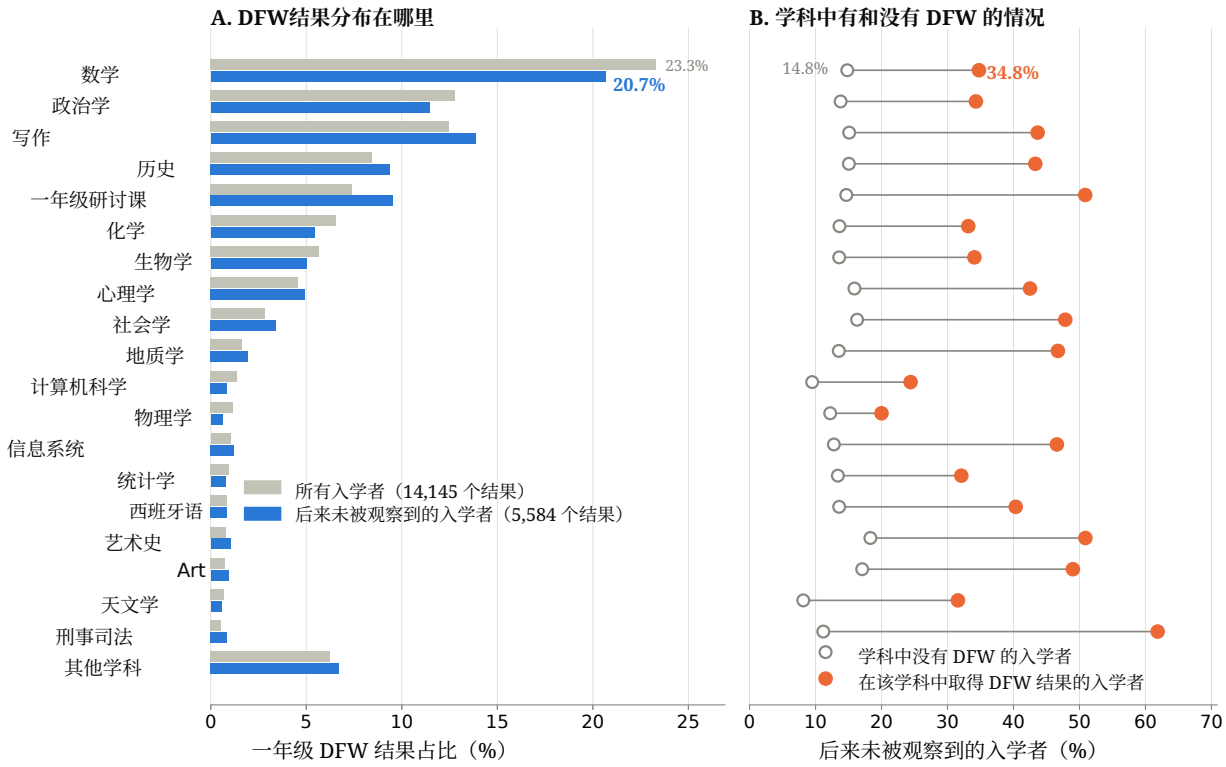
B.7 入门课程结果、持续就读与专业变更

本小节为第 3.4 节提供支持。

持续就读。 表B3按学科比较了这样两组入学者中后来未被观察到的比例：一组是在第一学年首次修读某门入门课程时取得 D、F、退课或无学分结果（DFW）的入学者；另一组是修读了此类课程但未取得上述结果的入学者。分析使用参考队列中未带有协调招生计划标记变量的 8,837 名入学者。19 个学科的差异介于 7.8 至 50.7 个百分点之间，且所有区间均高于零。在第一学年至少修读过一门入门课程的 8,823 名入学者中，5,267 人（59.7%）至少有一门课程取得 DFW，其中 29.5 % 后来未被观察到；其他入学者中这一比例为 12.7%。

按课程逐一统计，14,145 个 DFW 结果中有 3,298 个属于数学课程。后来未被观察到的入学者占有 5,584 个结果，其中 1,154 个属于数学课程；在后来被观察到的入学者中，数学课程所占比例为 25.0%，其中 54.4% 至少有一个 DFW 结果。1,542 名入学者有四个或更多 DFW 结果。

在数学课程中，2,413 名在已完成课程中取得 D 或 F 的入学者，之后不再出现的比例为 32.5 %（与没有 DFW 的入学者相差 17.7 个百分点，区间为 15.5 至 19.8）；751 名退课或取得无学分结果的入学者中，这一比例仅在学期结束后升至 42.1%。退课可能是已经开始的离校过程的一部分。在有数学 DFW 的入学者中，15.1% 在该课程所在学期之后没有记录。在这些入学者中，2,490 人在另一门入门学科中也有 DFW，674 人则没有。前一组与没有数学



图B6：参考队列中未带有协调招生计划标记变量的 8,837 名入学者，按学科划分的一年级 DFW 结果和持续就读情况。DFW 结果指第一学年首次修读一门入门课程，最终成绩为 D、F、退课或无学分结果。后来未被观察到：在入学学期一年或更久之后的学期中没有课程记录。(A) 所有入学者（灰色，14,145 个结果）和后来未被观察到的入学者（蓝色，由 1,553 名入学者取得的 5,584 个结果）中，各学科 DFW 结果所占的比例。(B) 在该学科中取得 DFW 结果的入学者（橙色）以及修读该学科但未取得 DFW 结果的入学者（空心圆）中，后来未被观察到的入学者所占比例。入学者至少为 400 人的学科，按其在所有 DFW 结果中所占比例排序；其他学科合并列入 A 面板。比较未经调整。

DFW 的入学者相差 21.6 个百分点（18.9 至 24.3），后一组相差 1.6 个百分点（−1.3 至 4.5）。纳入协调招生计划的入学者后，数学方面的差距为 11.3 个百分点（9.3 至 13.2），因为无论成绩如何，这些入学者都会在第一学年结束后离校。

有三项限制。比较结果未根据准备程度、第一专业和课程负荷进行调整。记录无法区分离开高等教育与转学到另一所院校的情况。此外，入学者可能在多个学科中都出现 DFW，因此各行描述的对象彼此重叠，不能相加。

专业变更。 表 B4 按学科比较更换专业的入学者比例：一组是在第一学年首次修读入门课程时取得 D、F、退课或无学分结果（DFW）的入学者，另一组是修读此类课程但未出现上述结果的入学者。表中列出两个总体，因为比较结果取决于哪些入学者仍可被观察到。

第一个总体是参考队列中观察到在入学至少一年后仍注册就读的7,250名入学者：他们在入学学期一年或更晚的某个学期有注册记录。曾暂时停学后又返校的入学者也符合条件，因此这一规则比注册就读

表B3：参考队列中按学科划分的入门课程第一年结果和持续就读情况。对于每个学科，将第一学年修读过编号低于2000的课程且有首次修读记录的入学者，按是否有任何一次此类首次修读以D、F、退课或无学分结果（DFW）结束进行分组。之后未观察到：从入学学期一年或更晚起，直至观察窗口结束（第2.2节），没有课程记录的入学者；曾暂时停学后又返校的入学者计为观察到，而是否转入其他院校无法与离开高等教育区分。该表使用队列中未带协调招生计划标记变量的8,837名入学者（总计10,027名）；带有该标记变量的1,190名入学者按该计划的设计会在第一学年后转学。括号内为逐点95%区间：差值使用Wald区间，单位为百分点；优势比使用对数（Woolf）法。列出有至少400名入学者修读过此类课程的学科，并按出现DFW的入学者人数排序。每名入学者可能出现在多行中。比较未经调整，且属探索性分析：这些比较是在分析计划冻结后计算的，区间也未针对学科数量进行调整。

			之后未观察到（%）		差异	
学科	入学者	有 DFW	有 DFW	没有	（点数）	优势比
数学	7,782	3,164	34.8	14.8	+20.0(18.0, 21.9)	3.07 (2.75，3.42)
政治学	5,050	1,673	34.3	13.8	+20.5(18。0，23。1)	3.26 (2.83, 3.75)
写作	6,036	1,665	43.7	15.1	+28.6(26.0, 31.2)	4.36 (3.84,4.95)
历史	4,829	1,076	43.3	15.1	+28.3(25.1, 31.4)	4.31 (3.71, 5.01)
一年级研讨课	7,995	1,046	50.9	14.7	+36.2 (33.1,39.3)	6.02（5.25,6.92）
化学	3,554	890	33.1	13.6	+19.5(16.2, 22.9)	3.14（2.63, 3.76）
生物学	2,667	760	34.1	13.6	+20.5(16.8, 24.2)	3.29 (2.70，4.01)
心理学	2,585	647	42.5	15.9	+26.6 (22。5，30。8)	3.91 (3.21, 4.77)
社会学	2,125	395	47.8	16.3	+31.5 (26.3,36.8)	4.71 (3.72, 5.96)
地质学	1,094	229	46.7	13.5	+33.2(26.3, 40.1)	5.61（4.05,7.76）
计算机科学	600	168	24.4	9.5	+14.9(7.9, 22.0)	3.08 (1.91,4.96)
信息系统	1,359	146	46.6	12.8	+33.8 (25.5,42.1)	5.95 (4.13,8.58)
物理学	554	145	20.0	12.2	+7.8 (0.5,15.0)	1.79 (1.09, 2.97)
统计学	881	134	32.1	13.4	+18.7(10.4, 27.0)	3.06（2.01,4.65）
西班牙语	627	119	40.3	13.6	+26.8 (17.4,36.1)	4.30 (2.75,6.72)
艺术史	527	112	50.9	18.3	+32.6(22.6, 42.6)	4.62（2.96，7.22）
Art	797	100	49.0	17.1	+31.9(21.7, 42.1)	4.67（3.01, 7.24）
天文学	402	95	31.6	8.1	+23.4 (13.6, 33.3)	5.21 (2.87，9.44)
刑事司法	407	76	61.8	11.2	+50.7 (39。2，62。1)	12.88（7.24,22.89）

截至一周年时。在这一总体中，56.5% 的人更换了专业。第二个总体是所有入学者，无论他们之后是否继续入学。首学年后没有记录的 2,777 名入学者，记录专业变更的机会较少，其中 7.7% 的人更换了专业。是否继续保持可观察状态本身可能取决于成绩、学业计划、是否通过协调招生计划转学以及是否离校，因此，第一个总体是经过筛选的，其比较结果仅描述后来仍可观察到的入学者。

数学方面，在后续可观察到的入学者中，差异为 10.6 个百分点（区间为 8.1 至 13.1；优势比为 1.56，区间为 1.40 至 1.73）；在所有有第一学年修读记录的 8,892 名入学者中，差异为 2.8 个点数（区间为 0.7 至 5.0；优势比为 1.12，区间为 1.03 至 1.22）。化学方面，这两个差异分别为 12.8 个点数（区间为 8.7 至 17.0）和 0.3 个点数（区间为 -3.4 至 4.1）。写作、历史和一年级研讨课中，所有入学者的差异均为负值（分别为 -10.2、-9.6 和 -16.0 个点数）。数学和物理学是仅有的在两个人群中差异均为正值且区间高于零的学科。首次记录的专业为大学学院的入学者尚未申报专业，并会在申报专业时更换专业。排除这些入学者后，后续可观察到的入学者中的数学差异

表 B4：参考队列中按学科划分的入门课程第一年结果与后续专业状态。对于每门学科，将第一学年内首次修读编号低于 2000 的课程的入学者，按这次首次修读是否以 D、F、退课或无学分结果（DFW）告终进行分类。更换专业：观察窗口内首次和最后一次记录的专业不同（第 2.2 节）。第 2 至第 7 列采用入学后至少一年仍可观察到在读记录的 7,250 名入学者（即在入学学期一年或更晚的某个学期有记录；暂时停学后返校的入学者也符合条件）。末列给出所有有此类首次修读记录的入学者中的相同差异（该队列共有 10,027 名入学者），不论其后续入学情况如何。括号内为 95% 区间：差异（以百分点计）采用 Wald区间，优势比采用对数（Woolf）法。列出后续可观察到的入学者至少有 400 人的学科，并按出现 DFW 的人数排序。同一名入学者可能出现在多行中。这些比较未经调整，仅作描述；它们是在分析计划冻结后计算的，不能确定课程结果与专业变更的先后顺序。

	观察到入学至少一年后仍在读						所有入学者
			更换专业（%）		差异		差异
学科	有 DFW 的入学者	无 DFW 的入学者			（点数）	优势比	（点数）
数学	6,390	2,196	65.2	54.6	+10.6(8.1, 13.1)	1.56 (1.40,1.73)	+2.8 (0.7,5.0)
政治学	4,301	1,133	60.1	57.2	+2.9 （−0.5，6。2）	1.13 (0.98, 1.29)	−2.0(−4.7，0。8)
写作	4,937	976	57.0	58.4	−1.4 (−4.9，2。0)	0.94 (0.82, 1.09)	−10.2 (−12.9，−7.6)
历史	4,020	634	58.8	59.2	−0.4 (−4.6，3。8)	0.98 (0.83，1.17)	−9.6 (−12.8，−6.4)
化学	2,982	605	69.3	56.4	+12.8(8.7, 17.0)	1.74 (1.44,2.11)	+0.3 （−3.4，4。1）
生物学	2,335	543	63.5	55.9	+7.6(3。0，12。3)	1.37 (1.13,1.67)	+1.7 (−2.2,5.7)
一年级研讨课	6,846	531	52.4	57.2	−4.8 (−9.2, −0.4)	0.82 (0.69,0.98)	−16.0 (−19.0, −13.1)
心理学	2,174	395	60.3	55.5	+4.8 （−0.6，10。1）	1.22 (0.97,1.52)	−3.9 (−8.1,0.3)
社会学	1,803	219	57.5	57.9	−0.4 (−7.3，6。6)	0.99 (0.74, 1.31)	−7.9 (−13.0，−2.9)
计算机科学	534	130	63.8	46.5	+17.3(7.7, 26.9)	2.03 (1.35,3.05)	+7.3 (−1.5，16。0)
物理学	550	127	55.1	44.4	+10.7 (0.8,20.5)	1.54 (1.03，2.29)	+12.9(4.5, 21.3)
地质学	901	124	58.1	55.2	+2.9 (−6.5，12。2)	1.12 (0.77, 1.65)	−11.0 (−18.0，−4.1)
统计学	783	97	60.8	51.0	+9.8(−0.6，20。2)	1.49（0.97，2.30）	+2.3(−6.4，11。1)
西班牙语	606	79	64.6	63.6	+1.0(−10.3，12。3)	1.04 (0。64，1。71)	−3.2(−12.4，5。9)
信息系统	1,162	78	75.6	64.9	+10.7 (0.8,20.6)	1.68 (0.98, 2.85)	−11.0 (−19.5，−2.4)
艺术史	428	60	46.7	55.2	−8.5 (−22.1,5.1)	0.71（0.41, 1.23）	−16.1 (−24.9，−7.4)
Art	644	52	59.6	57.6	+2.0 (−11.9，15。9)	1.09 (0。61，1。94)	−15.5 (−25.5，−5.5)

在 5,275 名入学者中为 13.2 个百分点（优势比 1.70，1.52 至 1.91）。

专业变更的标记变量遵循队列表的规则（第2.2节）：入学学期之前的记录也予以纳入（34 名学生有此类记录），缺失首次或最后一个专业也计为更换专业（27 名学生）。队列中没有学生在观察窗口的首学期或末学期同时拥有两个专业，因此同一学期内记录的顺序不会影响标记变量。根据仅保留已确认变更的入学界定规则，在两个专业均有记录的 10,000 名学生中，42.7% 更换了专业（纳入入学学期之前的记录时为 42.8%，如第2.2节所述）；后续观察到的入学者数学成绩差异为 10.7 点（8.2 至 13.2），所有入学者的差异为 2.9 点（0.7 至 5.0）。

持续性比较的两项局限在此同样适用：比较未经调整，且各行之间存在重叠。另有两项局限是专业变更特有的。专业变更是根据首次和最后一次记录的专业来衡量的，因此无法确定其发生时间与课程结果之间的先后关系。此外，比较了十七个学科，却未对多重性进行调整。这项分析属于描述性分析，且是在分析计划冻结后进行的。

B.8 班级规模：完整结果与敏感性检验

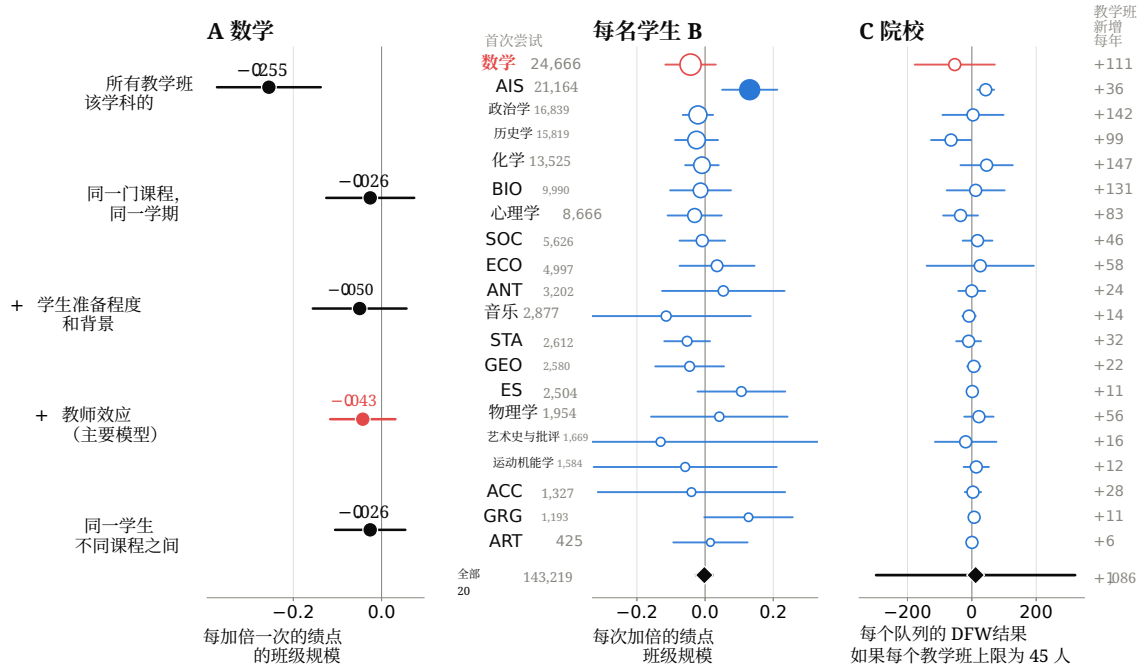
本小节为第3.6节提供支持，该节陈述了数学方面的结果。本文报告了跨模型和学科的比较（图B7）、情景分析（表B5）以及敏感性检验。

从所有教学班到主要模型。 图B7的A面板展示了随着比较范围逐步收窄，数学系数的变化。在所有教学班中，班级规模每加倍，绩点下降 -0.255点；纳入课程与学期效应后，降幅为 -0.026；而在主要模型中，降幅为 -0.043。相邻模型之间的差异是描述性的，不能准确地视为分别由课程、学生或授课教师导致的占比。将每名学生与其在其他课程中的自身表现进行比较，结果为 -0.026。在同一数学课程和学期内，班级规模每加倍，入学数学成绩相差 $+0.007$ 个标准差。测得的这一成绩指标上的平衡情况与按教室可用情况分配相符；但这并不能证明分配近似随机，课表、课程形式、专业或重修状态方面仍可能存在差异。

学科。 20个学科的合并区间（第3.6节）纳入了学科估计值之间的协方差，而此处的协方差接近于零。一个学科通过了错误发现控制，且方向相反：在一年级研讨课（AIS）中，较大的教学班与较高的成绩相关（$+0.131$）。该课程的班级规模中位数为25名学生，第90百分位数为50名学生，这表明在该课程中，班级规模可能反映了课程形式。

系数表示每名学生对应的关联，而各学科所服务的入学者人数差异很大：数学有24,666次首次尝试，艺术有425次。因此，图B7的B面板按首次尝试次数排列各学科，并根据该数量调整每个标记的大小；C面板则以院校的尺度呈现各学科：根据表B5的情景，预计每个入学队列的DFW结果会发生怎样的变化。该情景将该学科的DFW系数乘以受影响的首次尝试次数，再乘以班级规模的变化。按这一尺度，预计降幅最大的是历史学（每个队列的DFW结果为-65 -127至 -3）和数学（-53–178 至 $+71$）；预计增幅最大的是化学（$+46$ -35 至 $+127$）和一年级研讨课（$+43$ $+17$ 至 $+69$）。如果入学者人数很少，那么即使系数很大，影响也有限：音乐的每名学生负系数排名第二（-0.114），但预计每个队列的DFW结果仅变化 -9 （-29 至 $+12$）。没有任何学科的DFW系数通过错误发现控制。像同一学生模型那样，对每次首次尝试赋予相同权重后，每加倍一次对应-0.008 绩点（-0.031至 $+0.015$）。

情景。表B5报告了这样一种情景：将所有学生人数超过45人的讲座班拆分为人数相等且每班不超过45人的班级。该计算将拟合的规模关联应用于拟议的规模缩减，同时保持人员配置及其他条件不变（第 2.4 节说明了在何种条件下，这一计算可用于描述干预）。在数学学科中，受影响的首次修读所在班级的学生人数中位数为104人；拆分后，这些学生所在班级的学生人数中位数将为37人。其中，85%的首次修读发生在同时开设了学生人数不超过45人的班级的课程和学期中。这表明这些课程开设中确实存在规模较小的班级，但不能据此认定不同班级规模的学生和授课教师具有可比性。每个入学队列中，受影响的首次修读有3,708次，占数学学科一年级首次修读的三分之二。对这些修读整体而言，预测每个队列的DFW结果将减少53例，区间为减少178例至增加71例。在全部20个学科中，采用同一阈值将使讲座班每年增加1,086个（目前每年开设1,086个），预测每个入学队列的DFW结果变化为 $+12$ （-297至 $+321$）。



图B7：2014年秋季至2018年秋季秋季首次入学者在学生人数超过10人的面对面授课班级中的班级规模与首次修课成绩结果。区间为95%区间，来自对授课教师进行的野生聚类自助法（同一学生这一行采用按授课教师聚类的标准误）。(A) 数学（24,666次首次修读、18,878名学生、1,040个教学班、73名授课教师）：随着比较范围逐步缩小，班级规模加倍所对应的绩点差异。该学科所有教学班的系数为 -0.255；加入课程与学期效应、学生协变量及教师效应后（预先指定的主要模型，红色），系数为 -0.043（-0.117 至$+0.031$；$p = 0.258$），相应的DFW 比率差异为 $+1.0$ 个百分点（-1.3至 $+3.3$）。最后一行将每名学生与其在其他课程中的表现进行比较。(B) 20个学科各自的主要系数，按其服务的首次修读次数排序（代码旁标出次数，从MAT的24,666次到ART的425次），标记面积与该次数成正比；实心标记表示通过错误发现控制，阈值为 $q = 0.05$（20个中有1个）。菱形表示随机效应合并估计值，-0.002（-0.026 至 $+0.022$），其区间纳入了学科估计值之间的协方差。坐标轴根据区间跨度的第90百分位数设定；有2个较宽的区间超出范围。(C) 同一批学科按院校尺度呈现：若将所有学生人数超过45人的讲座班拆分（表B5所述情景），每个入学队列的DFW结果预计如何变化；计算方式是将各学科的DFW系数乘以受影响的首次修读次数，再乘以其班级规模变化；每年新增的讲座班数量标于右侧。数学：-53（-178 至 $+71$）；全部20名受试者（菱形）：$+12$（-297 至 $+321$）。没有任何学科的DFW系数通过错误发现控制。所有估计值均为以入学为条件的关联。课程班分配由行政安排；只有当关联确实反映规模的影响时，C面板的预测才适用。学科代码：MAT，数学；AIS，一年级研讨课；POL，政治学；HIS，历史学；CHE，化学；BIO，生物学；PSY，心理学；SOC，社会学；ECO，经济学；ANT，人类学；MUS，音乐；STA，统计学；GEO，地质学；ES，环境科学；PHY，物理学；AHC，艺术史与批评；KIN，运动机能学；ACC，会计学；GRG，地理学；ART，艺术。

表B5：情景：将所有超过45名学生的面授讲座班拆分为人数均等且每班不超过45人的教学班。预测变化将各学科的主要课程内系数应用于受影响教学班中秋季首次入学者的首次修读（2014年秋季至2018年秋季入学队列的年度均值），区间为根据该系数的野生聚类自助法计算得出的95% 区间。每年班级数统计该学科所有低年级讲座班。 †按入学数学成绩三分位划分的行使用包含规模与成绩交互项的模型；每一行都是规模系数与交互项的线性对比，并附有各自的野生聚类自助法区间，该区间是在方案确定后计算的，以替代方案所规定的聚类-t 区间。只有在班级规模导致估计差异、拆分范围内的关联关于规模的对数呈线性，且新增教学班的授课方式与现有教学班相同的情况下，该预测才描述一项干预；观察性模型无法确立这些条件。总计的区间包含各学科估计值之间的协方差。45名学生这一阈值属于情景设定值：记录中没有招生上限字段。

	首次尝试 每个队列	成绩 点数	DFW 比率， 百分比点数	DFW结果 每年班级数 每队列	now	增加
数学	3,708 +0.06 (−0.05 至 +0.17)	−1.4 （−4.8至 +1.9）	−53 (−178至 +71)	246		111
最低三分位†	1、181 +0.12 （−0.01至 +0.25）	−4.6 （−8.9至 −0.4）	−55 （−105至 −5）			
中间三分位†	1、409 +0.06 （−0.05至 +0.17）	−1.6 （−5.1至 +2.0）	−22 （−72 至 +28）			
最高三分位†	1，111 −0.01 (−0.11至 +0.10)	+2.0 (−1.3至 +5.2)	+22 (−14 至 +58)			
全部20名受试者	23,622		+12 （−297至 +321）	1,086		1,086

情景的区间来自对每项对比进行的野生聚类自助法。对于入学分数最低的三分之一数学首次修读，预测每100次受影响的首次修读中DFW结果减少4.6个，其区间为 -8.9 至 -0.4；反演自助法检验得到的区间为 -8.9至 -0.2，而方案规定的聚类-t 区间较窄（-8.2至 -1.0）。

敏感性检验。 有两项敏感性检验与数学结果有关。第一项涉及评分标准。入学秋季学期的19,271次数学首次修读中，有7,968次在一年内接续修读了另一门数学课程，其中6,510次在两门课程中都有评分记录。在这些继续修读的学生中，在给定首次成绩的情况下，后续课程成绩的系数为首次课程班级规模每加倍一次增加 $+0.027$ 绩点，区间包含零（-0.077至 $+0.131$）。其中5,934名学生的后续课程由不同教师授课，在这一子样本中，系数为 $+0.026$（-0.083 至 $+0.135$）。这项检验使用的是选定样本，即继续修读的入学者，并控制了首次成绩。检验未发现明确的关联证据，也不能排除小班与大班之间评分标准或学习效果存在差异。第二项检验涉及函数形式：各规模组的系数按顺序递减（11至30名学生为$+0.055$ ，46至75名为$+0.007$ ，76至120名为 -0.058，120名以上为 -0.085 ；参照组为31至45人的教学班），且没有一个系数显著异于零。还有一个数值仅具描述性，因为它是在方案定稿前观察到的：第一学期数学班级规模较大的入学者，一年内继续修读另一门数学课程的可能性较低（规模每加倍一次降低-2.95个百分点，区间为 -4.9至 -1.0）。

B.9 参考队列进展

表 B6 报告了参考队列的学业进展数据。在首个学期介于 2014年秋季至 2015年秋季之间的 10,027 名首次入学新生中，50.5% 在观察窗口内修读了高年级课程（即有注册记录显示其修读了编号高于 4000 的课程；观察窗口为入学学期至四年后的同一学期，第2.2节），43.0% 更换了专业，

表 B6：参考队列的基线数据：2014年秋季至 2015年秋季入学的首次入学新生；结果在所定义的观察窗口内测量（从入学学期至四年后的同一学期，含首尾；第2.2节），因此所有测量均在2020年秋季之前结束。由 cohort tables.py 根据机构数据库生成；没有任何内容是手工抄录的。

数量	N	占队列的比例
2014年秋季至2015年秋季入学的首次入学新生	10,027	100.0%
在观察窗口内修读了编号高于 4000 的课程	5,062	50.5%
在观察窗口内更换专业	4,311	43.0%
在观察窗口内修读数学课程	9,126	91.0%
在观察窗口内修读过 WRC	7,058	70.4%
在观察窗口内修读过 AIS	9,283	92.6%

同一观察窗口内。在观察窗口内修读入门课程的比例几乎普遍适用于 MAT 和 AIS（分别为 91.0% 和 92.6% ），WRC 的比例也很高（ 70.4% ）。各门入门课程与学业进展之间的队列限定关联见表 B2（附录 B.6）：修读 MAT 和 WRC 与观察窗口内未注册高年级课程的优势较高相关（优势比分别为 1.46 和 1.39）；而在设计限定为入学队列后，AIS 则转为强烈的有利关联（0.23）。结果指标是注册记录；没有此类记录的学生可能已经转学，正如 CAP入学者计划的那样，因此该结果衡量的是学生在本院校的学业进展，并不能衡量其是否从高等教育中流失。三门课程均有一项注意事项：在观察期内修读某门课程，意味着学生必须坚持就读足够长的时间，因此暴露组更倾向于由留存下来的学生构成；此外，两组在准备程度、学分要求、专业和修读时间方面也存在差异。净偏差的方向未知。此处的暴露指在观察窗口内任何时间修读过课程代码属于该学科的课程；队列中包括按设计会转学的 CAP入学者（第 2.2节）；未带 CAP标记的学生中，数学优势比为 1.42。1.46 的优势比对应的未进阶比例为：修读过数学课程的学生为 50.4%，未修读者为 41.1%。这些是选定入学队列中的关联。附录 B.4 的关联模型直接根据准备程度和背景衡量第一年表现。

B.10 基线参考量

基线的主要量汇总如下。每项均由基于哈希验证来源的可复现流程生成，且均注明其总体和数据年份；这些量共同构成2020年前的参照状态，可据此衡量中断后的比较。

1. **队列进阶**。 在首个学期介于2014年秋季和2015年秋季之间的 $N = 10{,}027$名首次入学新生中，50.5%在观察窗口内（从入学学期到四年后的同一学期，第2.2节）修读了高年级课程，43.0%更换了专业；门户课程的暴露率分别为：数学（MAT）91.0%、WRC 70.4%、AIS 92.6%（表 B6）。

2. **门户课程关联**。 队列限制优势比（表示在观察窗口内未注册高年级课程）：在观察窗口任意时点接触相关课程时，MAT 为1.46、WRC 为1.39、AIS 为0.23；需注意附录B.9中的限定条件（表 B2）。

3. **相对-gr　　成绩结构**。按学科和初始专业划分的成绩偏差矩阵　　r

在队列的专业变更者中。低于专业变更学生的常模：每位专业变更学生的化学（-0.50 GPA 分数），其次是数学和政治学（各-0.43 ），这两门学科在所有专业变更学生中的总偏差最大（-1，662 和 -1，593 GPA 分数）。高于专业变更学生的常模：每位专业变更学生的信息系统（$+0.61$），以及总计高于常模的一年级研讨课（$+1$，487）（图4）。

4. **全国排名**。 2019–20 版 College Scorecard 数据中的坐标：本科生人数为 27,573，标准修业时间 150 % 内的完成率为 0.423，佩尔助学金占比为 0.440，相对于以佩尔助学金占比为条件的全国预期值的残差为 -0.055；当前 2024–25 版数据中的坐标：本科生人数为 30,580，完成率为 0.526，佩尔助学金占比为 0.441，残差为 $+0.064$（图 B5）。

5. **学区与生源背景**。 在 15 个课程类别中，按各课程相应年级的估算学生人数计算课程选修人数：代数 II、几何和生物学与学区儿童贫困的关联接近于零；微积分以及 AP 数学与科学的课程选修人数每当儿童贫困率上升 10 个百分点，便下降相当于全国平均值 36% 至 37% 的幅度；7至8年级代数 I 的点估计为 16.4%，但主周期的区间包含零；微积分斜率（每单位贫困率为 -0.171 ，按州聚类的标准误为 0.020）及其与学区收入的相关系数（0.45）；在贝克萨尔县的 15 个学区中，同样的对比显示：贫困程度最低的两个学区组中，微积分选修率分别为全国比率的 1.02 倍和 0.92 倍，其他三个组则为 0.38 至 0.45 倍，县级梯度为 -32%（-63 至 -11）；生源学校课程供给概况及其方向相反的学分入学占比，以及生源学区贫困分布（按学生加权的中位数为 16.3%，而得克萨斯州按儿童加权的中位数为 21.9%）（图 1、2、B1 和 B2）。

6. **准备程度与第一年表现**。在2014年秋季至2019年秋季的23,489名关联入学者中——这一总体是在该方案首次应用于11个队列后选定的——入学数学成绩的变异有22%来自生源学校和学区之间的差异；学校高阶数学指数每增加一个标准差，成绩相差 $+9.7$ 点；学区儿童贫困程度每增加一个标准差，成绩相差 -10.4 ；在控制语言成绩后，数学成绩每增加一个标准差，第一学年非定量学科成绩提高0.140点（0.109至0.171），定量学科成绩提高0.263点（按首次尝试加权的学科中，分别为0.125和0.316）；排除生源学校后，在纳入其他模块的情况下，学区贫困程度和学校模块会使相对成绩平方误差的比例降幅中位数分别变化 $+0.0004$ 和 -0.0015 （表 B1、图 B4 和图 B3）。

7. **班级规模**。 在2014年秋季至2018年秋季入学者的数学面对面讲座班首次修读中，对于入学数学成绩较低的入学者，较大的班级规模与更多的DFW结果相关，这为检验较小班级规模提出了一项假设：班级规模与入学数学成绩的交互项为每个标准差 $+0.046$绩点（0.005至0.088）；对于入学成绩最低的三分之一学生，45人情景预计每100次受影响的首次修读中DFW结果减少4.6个（-8.9 至 -0.4），绩点增加0.12（-0.01 至 $+0.25$）。对所有入学者取平均后，在控制课程-学期和教师效应的情况下，班级规模每加倍，关联为 -0.043 绩点（-0.117至 $+0.031$）；相比之下，该学科所有教学班的关联为 -0.255 ，20个学科合并后的关联为-0.002（图B7，表 B5）。

支撑这些数量的院校课程结果，均记录于2020年春季中断事件之前。三项内容使用了较晚或未注明日期的来源，并据此加以标注：学区分析包含一个2021–22稳健性周期，全国排名包含2024–25年记分卡

数据年份，而提取记录未注明助学金变量、入学测试或每名学生的学分表的提取日期。表 C1列出了每项分析的观察窗口和版本。每个队列数量均使用第2.2节的观察窗口，包括上边界；比较研究必须完全复现该边界。生成路径与图和表的片段一同记录；源文件通过数据清单中的 SHA-256 哈希值以及分析数据库的数据溯源层（第 2.1 节、附录 A.1）保存。

C 数据附录：材料、关联验证与源数据质量

本附录记录论文测量所依据的材料：各数据集及其版本；分析数据库及其溯源层；从入学者关联至其就读的高中，再关联至学区贫困程度和社区背景的构建与验证；推断惯例；选择审计；以及处理学生记录时所遵循的披露控制制度。本附录还记录了会危及分析可复现性的源数据缺陷，并列出持续适用的补救措施。其中一些缺陷会产生看似合理但实际错误的数值，却不会触发任何错误或警告；另一些缺陷存在于公开数据集（CRDC、SAIPE）中，因此任何将这些数据集关联起来的研究团队都会受到影响。第 C.3 节逐一说明各项缺陷、未处理时造成的后果，以及消除其影响的规则。表 C1 列出了论文中每项分析所对应的总体、单位和观察窗口。

C.1 数据来源与数据溯源

表 C2 列出了分析数据库中的所有数据集。该数据库（GATEWAY，基于 PostgreSQL 16.2）按溯源层分层：原始层逐字保存源文件及其文件名和 SHA-256 哈希值，且绝不编辑；暂存层为数据指定类型并隔离哨兵代码；标准化核心层包含维度表和事实表；分析就绪数据集市则为每幅图提供数据支持。每次原始数据加载都会登记在溯源表中（数据集、版本、路径、SHA-256、字节数、源网址、加载时间戳、行数），因此任何已发表的数值都可以追溯到相应文件及其哈希值。本论文直接获取的档案还包含完整的清单条目（标识符、机构、发布版本、获取网址和日期、许可证、路径、哈希值、大小、模式、解析器版本、沿袭关系、地理范围、时间覆盖范围、单位、隐私政策、修订政策）。全过程优先采用批量文件下载，而非 API 查询，因为文件可计算哈希值，也可离线复现。

有四项源数据层面的事实对解读至关重要。第一，SAIPE 学区序列覆盖 1995、1997、1999 年以及 2000–2024 年；1996 年和 1998 年从未发布过数据，而 1989 年和 1993 年的估算仅涉及县，因此学区贫困程度序列最早只能追溯到 1995 年。第二，五个 CRDC 周期分为四种不同的套件格式，使用两套不同的列名称体系；2011–12 周期在列名称体系、缺失值约定和学区报告率方面（高出约 40%）均不同于其后所有周期。纳入的五个周期均为总体集合：美国国家教育统计中心将 2011–12 年起的 CRDC 记录为总体集合（此前采用抽样的周期为 2004、2006 和 2009–10 年），因此这里各周期之间的差异，是不同总体集合之间覆盖范围和约定的差异。相关影响将在第 C.3 节中阐述。第三，CRDC 学年与 SAIPE 收入年份之间的日历时间对齐仅记录一次：在一张维度表中，将每个 CRDC 周期与其秋季年份对应的 SAIPE 数据版本配对（2017–18 $\rightarrow$ 2017）；所有连接均通过该表进行，而不是在每个查询中重新推导这一约定。第四，学区在三十年间不断变化：SAIPE 中至少出现过一次的学区标识符共有 15,007 个，其中只有 12,659 个（84.4%）在每个版本中都出现，因此存续期维度记录了每个标识符的首次和末次出现年份，并

表 C1：按分析列示的总体和观察窗口。每一行的院校课程结果均记录于 2020年春季中断之前；Scorecard 比较以及学区分析的一次稳健性周期使用了较晚的数据版本，并已明确标注。该提取数据不包含记录助学金申请、调整后总收入或入学测试日期的字段（第 2.4 节）。

分析（教学班）	总体和单位	观察窗口或数据版本
学区课程选修人数（3.1、3.2）	9,790个学区，以及其中的15个学区贝克萨尔县；学区CRDC课程注册人数总计和SAIPE儿童贫困	CRDC 2017–18与SAIPE 2017；基于民权数据收集（CRDC）2015–16年的稳健性检验，以及2021–22
生源学校学分概况和生源学区贫困状况（B.2、B.3）	来自 1,106 所生源学校的 48,168 名学生所学校的学生，以及来自 682 个得克萨斯州学区；按汇总计数统计学校层级或学区的2014年秋季至2019年秋季期间活跃的学生	CRDC 2017–18学校概况；平均值SAIPE 2010–2018年数据版本
入学准备程度（B.4）	23,489名关联的秋季首次入学者；每名学生一行	2014年秋季至2019年秋季的入学队列；来自最接近高中毕业年级的CRDC周期的学校背景最接近高中毕业年级
第一年表现（B.4）	19,874名已关联的秋季首次入学者；首先低年级课程的修读次数	2014年秋季至2018年秋季的入学队列；入学秋季学期以及随后春季和夏季
班级规模（3.6，模型）	23,766 名秋季首次入学者，已关联或未关联；面对面授课班级中的 177,289 次首次修读讲座班	2014年秋季至2018年秋季的入学队列；入学秋季学期及随后春季和夏季
参考队列（3.4，3.5、B.6、B.7、B.9）	10,027名首次入学新生，入学时间为秋季2014年至2015年秋季；每名学生一行	从入学学期到四年后同一学期（含首尾）2014年秋季至2019年秋季
完整窗口描述统计（3.4，曲线3.6）	所有在读学生，入学者和转学生皆然；学生或教学班	
全国排名（B.5）	以学士学位为主的第四篇章院校	College Scorecard 数据年份 2019–20和 2024–25

表 C2：数据清单。各行是载入分析数据库后的情况；学生微观数据（标记为†）采用就地查询，从不导入（C.6 节）。

来源	Unit	观察窗口／版本	留存规模
SAIPE 学区贫困程度 [55]	学区–年份	1995–2024，28 个数据版本	382,099 行
民权数据收集（CRDC）学校文件 [56]	学校–周期	5 个周期，2011–12 至2021–22	481,347 行
学区汇总	学区—周期	same	84,361 行
课程类别入学人数（15 个类别；7在2013–14年）	学校–周期×课程	2013–14、2015–16年，2017–18年、2021–22	5,030,159行
美国国家教育统计中心 EDGE 地理编码 [57]	学校–周期	2017–18年，2021–22	204,467 行
NCES CCD 名录 [58]	学校	2010–11，2013–14	仅用于恢复流程
机会地图集 [38]	人口普查分区	2010年人口普查区定义	74,123 个普查区
汇总至邮政编码制表区	ZCTA	2010年定义	28,744 个邮政编码制表区
邮政编码制表区与人口普查区关系 [49]	邮政编码制表区×普查区	2010年定义	125,330 个权重
NCES EDGE GRF 人口普查区与学区关系 [42]	LEA×普查区	GRF 2017（SY 2015–16边界，2010 年普查区）	112,698 条记录
德克萨斯大学圣安东尼奥分校记录†	学生×学期×课程	学期：2009年秋季–2019年秋季 1	,984,991 行
时间段内活跃的学生†	学生	2014年秋季至2019年秋季	68,558名学生
每名学生的学分表†	学生	（见正文）	其中有 61,954 人

秩统计量基于平衡面板计算。

院校记录存储在大学的 SQL Server 中，其设计本身采用非规范化结构（每名学生、每个学期、每个课程教学班各占一行，并重复记录学生属性）。提取的数据包含 92,769 名不同的学生，其中 68,558 名学生在 2014年秋季至 2019年秋季期间有课程记录；本文所有院校层面的计数均针对这些在校学生。关键信息的覆盖率很高：其中 98.6% 的学生记录了生源高中名称，87.6% 记录了得克萨斯州的高中，99.96% 记录了家庭邮政编码。另有一张学生级表记录了 61,954 名在校学生在基础数学课程中的大学前学分（包括大学先修课程、双重注册和大学水平考试项目（CLEP）学分）；其提取窗口未在源系统中得到独立证实，且其中缺少 6,604 名在校学生的情况已记录为未知，因此本文绝不交替使用这两张表。关联入学者模型（第 2.4 节）还从课程行表中读取助学金申请标记、助学金申请者的调整后总收入、最高入学数学和语言成绩，以及 CAP 标记；收入在模型中以队列内排名的形式纳入，绝不制成表格或导出，学生行也仅保存在内存中。其 AP成绩字段仅记录字面值 CR，而双重注册成绩字段则记录字母成绩；此处记录了这一不对称之处，因为它限制了未来对学分成绩的分析。我们没有重复使用该项目早期的数据库，因为无法还原其表格的生成方式。我们根据源文件重建了公共数据表，并记录了这些文件的哈希值。

C.2 链接构建与验证

每项情境测量都通过两个分支之一关联到学生。学区分支将学生自由文本中的高中名称解析为NCES学校标识符，其前七位数字即为学区标识符（LEAID），可与SAIPE贫困率进行精确连接；一旦学校解析完成，该分支中的任何环节都不再具有概率性。社区分支则将学生居住地的邮政编码解析为机会地图集中的背景信息。两个分支的失败方式不同，因此分别进行了验证。

名称解析。交叉映射表针对该数据提取中的1,663个不同的得克萨斯州高中名称–城市–邮政编码组合构建；其中1,583个组合对应60,045名登记有得克萨斯州高中名称的在校学生。我们使用公式（A1）的得分，将这些组合与EDGE 2017–18目录进行匹配。完整字符串相似度在此存在结构性缺陷，因为院校记录使用完整的尊称名称，而NCES记录使用简称（$Sandra\ Day\ O'ConnorHigh\ Sch$ 与$O'CONNOR\ H\ S$），因此该得分采用优先包含匹配。自动接受要求得分至少为0.72。在全州区块内，还要求其得分至少比次优匹配项高0.08；如果两所真实学校之间近乎平分，则交由人工复核；在城市和邮政编码区块内，则取消这一差值要求，因为区块本身已能消除歧义。之所以在此明确说明这一规则，是因为这意味着接受的匹配集合包含一些差值较小的匹配，而这些匹配的错误率尚未依据经过裁定的金标准进行测量。得分低于0.45的候选项将被完全剔除，依据的原则是：错误的学区标识符比缺失的学区标识符更糟。

表C3展示了分阶段的覆盖情况。三个验证环节塑造了最终设计，每个环节都用一条明示规则取代了朴素做法。（i）低阈值匹配会产生得分很高的错误匹配：一次仅按州进行阻断、分数阈值为0.80的恢复流程，将一所六年级学校、一个完整学区、一所九年级校区和一所另类教育中心接受为高中；该流程随后被撤销，并在分数阈值提高至0.90且扩展排除名单后重新运行。现在，该流程会撤回此前任何一次运行所提升的所有条目。现行规则是：即便名义上的分数阈值很高，不按城市进行阻断的名称匹配仍不可靠。（ii）私立学校属于结构性缺失：118个名称（2,261名学生，占交叉映射基数的3.8%；右侧面板中的2,058

表C3：2014年秋季至2019年秋季期间在校学生的关联验证：学校对照表的分阶段覆盖情况，以及学生级表中61,954名在校学生的最终处置情况。左侧面板中的百分比以拥有得克萨斯州高中名称的60,045名在校学生（交叉映射基数）为准；右侧面板则对学生级表中的学生进行划分，每名学生仅计一次。

对照表阶段（可用学生）		最终处置（所有学生）	
自动匹配	88.3%	匹配的	49,099
+人工裁定（39 行）	89.7%	得克萨斯州以外的高中	6,945
+历史版本恢复	89.8%	未解决	2,749
+CCD目录恢复	90.7%	私立（不在美国国家教育统计中心范围内）	2,058
		未记录学校	859
		审核未通过	244

人是在学生级表中统计的同一群体）按定义不属于美国国家教育统计中心的公立学校范围；将其计为匹配失败，会使错误率被高估相应比例。（iii）名称模式并非合适的高中分类器：仅凭名称判断“是否为高中”，会在不知不觉中漏掉初高中合并校区，而这类校区多位于农村地区且规模较小，因此并非随机缺失。之后，一次利用CCD年级范围的恢复流程显示，该流程找回的学生中有96%原本一直都能在目录中找到；正确的分类依据是年级范围（最高年级为9至12年级），并已将其记录为任何重建工作均须遵循的现行规则。由于重建会使已完成的人工裁定失效，因此在覆盖率达到90.7%时并未执行重建；相反，对剩余部分进行了审计（C.5节）。

随后，我们使用一个外部来源对交叉映射表进行了验证：机构数据库中的一张学校级表，其数据溯源已通过与本项目导入的 CRDC 数据进行精确匹配连接得到确认（该表中的代数 II 列与 CRDC 2017–18 在 19,961 所学校上完全匹配，匹配率为 100.0%；相比之下，与其他所有周期的匹配率均为 2.0–7.7%）。该表涵盖了在校学生中 1,192 所匹配学校的 94.7%，证实通过名称匹配分配的标识符对应于全国范围内的真实学校。这同时独立验证了交叉映射表和 CRDC 导入数据；但仍无法确定在名称相似的候选学校中是否选对了学校。

邻里定位。 家庭邮编分支在使用前经过了验证，因为它基于一个未经检验的假设：记录的第一个邮政编码是童年住址。我们比较了在校学生记录的家庭邮政编码与记录的高中邮政编码对应的位置；只报告百分位数和占比。两者间距离的第90百分位数为21.2英里，第95百分位数为208.5英里。在距离记录的高中超过100英里的学生中，63.6% 的家庭邮政编码位于圣安东尼奥。这一模式与学生迁入城市后登记本地地址的情况相符；单凭地理位置无法确定地址字段记录的时间。因此，我们不会将完整的距离分布解读为通勤距离，距离仅用作链接门槛。主要样本为距离不超过10英里的学生（按邮编质心测量，占学生的80.0%）；下文的点坐标复核证实，在其子样本中这一阈值对应的比例为83.5%，而严格的敏感性变体采用5英里。学区归属始终通过已确定的学校来判定：仅贝克萨尔县就有十五个学区，因此偏移几英里通常就会跨越学区边界。这项诊断的初始版本还错误测量了学校端，使用了邮政编码质心：包含44.4%学生的零英里峰值，在将学校作为点测量后降至0.3%，而阈值未作任何调整。这一峰值是由

邮政编码质心几何造成的假象。

将学生向下分配到人口普查分区的做法经过检验后被否决。按学生实际居住地加权后，众数分区分配大约有五分之四的情况是错误的（最能代表中位数学生的分区，仅占其邮政编码对应总人口的 22%；平均而言，每个学生的邮政编码涵盖 9.9 个分区）。未加权的邮政编码制表区层级统计看起来尚可（中位数为 0.742），因为它描述的是邮政编码；按学生加权后则不成立。因此，设计方向反转：将机会地图集的数值向上汇总到邮政编码制表区，计算其所含各普查区的总体加权平均值，所用数据来自 2010 年人口普查关系文件（选择该文件而非 HUD 交叉对照表，是因为它可下载并可计算哈希值，记录的是人口权重而非地址重叠情况，且基于 2010 年人口普查区定义，与地图集的数据版本完全一致）。学生覆盖率为 97.7%。由此得到的变量具有双重生态层面属性，且这一约束在模式本身中得到落实：背景表存放于以其层级命名的模式中，每一行都包含一个“是否为生态层面”的标记变量和一个分区数据版本列；表注释（可在联接时查询）明确说明，没有任何数值描述个体。

C.3 源数据缺陷及常设补救措施

源文件中有八项属性，若处理不当，足以改变结果。表 C4 总结了每项属性及其补救措施；后果最严重的五项将在下文介绍，并附上揭示这些问题的检查方法。检测方式本身也是一项方法学成果：每项缺陷都是在某项稳健性或一致性检验中浮现的，而这些检验原本另有目的；其中两项缺陷在最初选定的设定下无法察觉。因此，这些检查也是方法的一部分。

受损的学校标识符（2015–16年）。2015–16年的民权数据收集（CRDC）宽表文件中的学校标识键存在数值损坏，其形式与上游某处发生的电子表格或科学计数法强制转换相符；确切的处理路径尚未核实，已观察到的事实是键不唯一。按存储形式读取时，96,360 所学校中有 66,986 所被归并到重复键上，该周期的代数 II 选课人数也从 2,931,194 降至 807,667，损失达 72%，从而形成一个看似合理、却出现神秘低谷的面板。现在，每个周期都会根据组成部分（地方教育机构标识符和 SCHID ）重建学校标识键，并在加载时执行唯一性断言。

显式零值与跳过代码（2011–12）。在 2011–12 年，26.25% 的学校将代数 II 选课人数记录为显式零值；此后的每个周期，这一比例都在 0.01–0.41% 之间。这一差异与编码惯例的变化同时出现。2011–12 周期与后续周期的不同之处在于列名称体系、缺失值的标记方式（前者用空白表示，后者用保留代码表示），以及学区报告率高出约 40%。在后续周期中，未开设某门课程的学校会在该课程的入学人数项目中使用跳过代码（下文已对 2017–18 年的数据进行精确核实）；这与 2011–12 年的零值来自未开设该课程的学校这一解释相符，但尚未依据 2011–12 年的文档核实。本文使用的解析器会将空白和保留代码转换为缺失值，并在求和结果全部缺失时保留缺失状态，因此这些零值是归档源文件中的显式值。因此，将“报告非负代数 II 人数”作为指标，从 2013–14 年起几乎恒定不变，原因在于编码方式：未开设该课程的学校不会进入已报告人数的范围。该指标几乎不能说明有多少学校开设这门课程。这个问题是在稳定性分析中发现的：其首个版本报告该指标的跨周期一致率为 93–100%；与之并列的 Cohen’s $\kappa \approx 0$ 显示该指标退化，而基于入学率重新计算后，零值方面的差异便显现出来。研究结果表明存在协调问题：要比较各周期的开课情况，需要使用等效的开课标记，并对零值和跳过代码采用统一处理方式；本文未进行这项工作。由此带来的影响是：没有任何开课比较涵盖 2011–12 年，

表C4：可能危及可复现分析的源数据缺陷。“无声”是指缺陷会产生看似合理的输出，且没有错误或警告。

Flaw	未处理时的后果	补救措施
CRDC保留码（离散值，例如 -3、-5、-6、-8、-9、-11；特定周期）存储在计数列中	无声地：SUM() 将 -9 哨兵值纳入注册总数	哨兵值 $\rightarrow$空值，代码保留在平行列中；抑制状态保持可报告
CRDC 2015–16年以数值形式提供受损的学校标识符	未被察觉：96,360所学校中有66,986所折叠到重复的标识符上；该周期的入学人数就此消失	将标识符重构为每个周期的 LEAID(7) ‖ SCHID(5)；存储的键不可靠
2011–12年民权数据收集记录了显式入学人数列中的零值，而后续周期记录的是跳过代码	未明示：26.25%的学校有一项代数 II 的显式零值（相较于每个后续周期中的 0.01–0.41%），这是编码惯例上的差异，使得朴素的开课指标被解读为学校发生了变化	没有任何开课比较涵盖 2011–12 之前的零值和跳过代码经过协调；秩统计量应结合这一注意事项解读
七个学区的代数 II 选课学生多于儿童 5–17 岁（比例最高达 31.3）	静默：反转未加权全国梯度	要求 ≥ 200 儿童年龄为 5–17 岁（关于该预处理筛选的分母进行预处理筛选）
课程名称是截断的自由文本拼写不一致的文本	无提示：一门课程被拆分为最多四个；课程级单元格低估	在一个键上进行规范化任何操作之前的无标点键分组
SAIPE 2024年重新排列了列在机器名称下人口统计字段是标签字符串	静默：贫困被读作总体若按位置解析静默地将数值强制转换为 NaN值整列因此从输出	按名称匹配列；按位置禁止解析显式标签比较；否数值强制转换
学校邮政编码少存了一位数字（76 名学生）	一所休斯敦学校会被识别为新在邮政编码前补零后，新泽西州左侧	已标记；因距离而排除 gate

而涵盖该周期的跨周期排名统计，解读时须注意：覆盖四分之一学校的一组并列零值会扰乱排名。

数据框架不匹配比率。 每名常住儿童的代数 II 选课人数，是用民权数据收集（CRDC）的分子除以小区域收入与贫困估计（SAIPE）的分母得出的；在非常小的学区中，这两个数据框架并不一致：2017–18年有 7 个学区记录的代数 II 选课学生多于 5–17 岁儿童（最大比率为 31.3；违规学区的中位数儿童数为 35），这是因为某所学校服务的区域比其所属学区更广所致。这 7 个学区占 10,425 个学区中的一部分，使非加权全国回归的符号发生反转：$+0.030$ 纳入这些学区时，相比于 -0.037 排除这些学区并且 -0.049 高于现行的 200 名儿童下限时（在这三种拟合中，加权斜率均为 -0.052 ）。此前这一缺陷之所以未被发现，是因为已发表的设定按学区规模加权，这种做法从机制上压低了 35 名儿童学区的影响；直到最终运行非加权稳健性拟合时，该问题才显现出来。补救措施是对分母进行筛选（对比率进行截断会以因变量为条件），当儿童数下限设为 200 时，最大比率变得符合实际（0.68），且加权斜率与非加权斜率的差异不超过 0.003；在整个筛选阈值序列中，加权斜率保持不变（-0.0521 至 -0.0527）。

课程碎片化。院校记录中的课程名称是自由文本，截断至约三十个字符，并且在 2012–2019 年观察窗口内拼写各异（商科微积分代数有四种拼法；科学家与工程师代数另有四种）。按原始名称分组时，

一门课程会被拆成多个片段，而片段与实际规模较小的课程无法区分：第一版分析按片段的“数量”选取课程，因此一个低于披露最低标准的单元格进入了结果表。通过不含标点的小写键进行规范化后，计数得以纠正，且未改变任何微积分课程轨道的数据；所有下游分析均可完全复现。

哨兵代码。CRDC 会在计数列（-3、 -5、 -6、-8、 -9和 -11 ）中存储离散的保留代码（在 2017–18 年公共使用文件中，以及特定周期的变体； -11 表示数据质量抑制），而其发布的总计会将这些代码视为零。直接转换类型并求和，会在不知不觉中将-9 个哨兵值计入注册总数；仅 Algebra II 字段一项，每个周期就涉及约 75,000 所学校。暂存层会将每个哨兵值替换为 NULL，并在平行列中保留其代码，从而使抑制情况仍可报告：这一处理所促成的分解，将表面上的覆盖危机（仅有 $\sim$22%的学校报告 Algebra II）厘清为结构性缺失（小学和初中学校，代码为 -9），从而消除了对学区总和存在向下偏差的疑虑。对此解读的一个限制也一并记录在案：公共使用手册采用避免披露调整，以扰动已发布计数，但不会留下哨兵值；因此，缺少代码只能证明不存在带代码的抑制，而这些计数应视为受保护的已发布数值，而非精确统计值。

第 3.2 节的课程类别比较采用了平行列所支持的另一种解读。对于课程项目，代码 -9 表示该学校没有高中成绩，或该高中报告该学科的班级数为零：2017–18年，报告至少开设一个代数 II 班级的 21,419 所学校，恰好就是代数 II 入学人数为非负数的学校。在开设高中年级的学区内，跳过某个课程项目因此属于结构性零值，在学区总数中按零计入；其他所有代码仍视为缺失。由此，学区总数有三种状态：计数、结构性零值，或有些学校持有缺失代码的不完整总数。2017–18年，代数 II 样本框中的五个学区在某个课程类别中总数不完整（2015–16年为三个，2021–22 年为十三个）；这些学区不纳入共同样本框，而将其保留也不会改变任何结论。另有两组学区不适用这一解读，需通过样本框处理。2017–18年，有 65 个开设高中年级的学区（共 247,131 名 5–17 岁儿童）报告九门高中核心数学与科学课程均无入学人数，这一报告缺口不在代数 II 样本框之内；另有 144 个学区（共 450,005 名儿童）报告代数 II 入学人数为零，但其他核心课程的入学人数为正。中性样本框变体纳入第二组学区（9,934 个学区，而非 9,790 个），且不改变各课程类别之间的对比。

上述五种缺陷都不易察觉：每项分析都顺利完成，没有报错，得出的数值也看似合理。因此，揭示这类缺陷的检查应成为常规做法：即使主要估计值看起来稳定，也要执行所述的稳健性检验；除了计算原始一致率，还要计算诸如 κ 之类的一致性统计量；并相互检验比率的分子框架和分母框架。

C.4 推断实践

不确定性是在暴露发生变化的层级上计算的。无论单元格中包含多少学生，学区贫困程度都因学区而异，州政策则因州而异；忽视这一点会使学区梯度的标准误窄约五倍（在点估计相同的情况下，按州聚类的标准误为 0.0136，而朴素估计为 0.0027）。重抽样和聚类遵循相同的原则：课程类别比较在州这一教育政策发生变化的层级上进行重抽样；关联入学者模型则在生源学校这一学校背景发生变化的层级上进行聚类。跨学科汇总时会使用学科估计值之间的协方差，因为同一批学生和学校会参与多个学科模型

。班级规模模型将授课教师作为聚类单位，并对每个规模项以及规模项之间的每种对比采用野生聚类自助法；此前在打乱的成绩上进行的试运行表明，当授课教师聚类数量较少且大小不均时，聚类 t 检验的拒绝率过高。检验多个学科类别时，Benjamini–Hochberg 程序会将假发现率控制在规定水平（附录A）；每项报告的关联均注明其研究设计和观察窗口。

C.5 选择审计

三种选拔机制可能使关联结果产生偏差，我们分别对其进行了直接审计。

尚未解决的剩余样本。 在样本中，2,749 名其得克萨斯州高中无法通过任何方法匹配的学生，与 49,099 名已匹配学生有所不同：他们持有双重注册学分的比例更高（17.1% 对 12.6%），持有 AP 学分的比例更低（2.07% 对 2.87%），平均数学入学成绩较低（549 对 555；即 SAT 数学成绩或换算后的 ACT 数学成绩中的较高者，采用 200–800 分制），获得佩尔助学金的比例也更高（48.9% 对 46.6%）。若将学生视为相互独立的个体，这些对比相对于其标准误较大（依次为 $z = 6.09$、-2.82、-3.46 和 2.34）。同一所学校的学生共享该校的学分课程设置和社区环境，而未匹配学生来自 197 个学校名称。以生源学校为聚类单位后，相同对比的数值为 $z = 1.40$、-1.51、-1.05 和 0.72。因此，这些差异处于学校之间的变异范围内；审计无法确定匹配是否在这些指标上存在选择性。由此造成的学区关联结果偏差方向尚不明确：被排除学生的学区贫困程度无法观测，而边际结果相近（首门课程不及格率为 32.5% 对 30.6%；在独立学生假设下为 $z = 1.71$，在学校聚类下为 1.02）并不能确定他们的背景与结果如何联合分布。附录 B.4 中的关联模型通过两种方式处理这一问题：使用关联逆概率权重，以及比较关联样本与所有入学者的系数；两种方法均未改变结果。

无法匹配到得克萨斯州学校的学生。6,945 名学生记录的高中位于得克萨斯州以外，不在交叉映射表的覆盖范围内；另有 859 名学生未记录高中名称。这两组学生大多是转学生（分别占 71.0% 和 100.0%，而表中所有学生的这一比例为 42.8%），他们的大学前学分由第一所院校记录，因此在这里大多不可见：两组学生中，记录有 AP 学分的比例分别为 1.0% 和 0.0%，记录有双重注册学分的比例分别为 3.0% 和 0.7%；表中所有学生的相应比例为 2.5% 和 11.5%。他们不纳入任何关联分析，绝不能与“无学分”学生合并统计。

私立学校差额。2,058 名学生就读于美国国家教育统计中心公立学校名录之外的学校。他们的条件明显更优（佩尔助学金比例为 25.7%，第一代大学生比例为 27.8%；匹配学生的相应比例为 46.6% 和 47.7%），因此将他们排除在外，会使每项学区关联结果都缺少一个条件更优的群体；其对任何梯度的影响尚未测量。

C.6 治理与披露控制

学生层级记录按照伦理审批声明中所列的机构审查委员会方案和数据申请进行处理，并遵循一项比正式要求更严格的治理规则：可以在本地查询和处理单个学生记录，但绝不将其摄取到分析数据库中。计算期间，学生行仅存在于内存中；对于较早的制图程序，则存放在本地缓存中。进入数据库的只有汇总单元格，每个单元格至少由十名学生支撑；这些数据通过一条代码路径发布，该路径会拒绝较小的单元格，并记录被抑制和已发布的单元格数量。

在该代码路径之外生成的文件输出和旧版表格，一经发现便会经过检查，并纳入同一最小单元格规则；自动扫描器会逐一列出所有计数单元格低于十的图表配套导出文件。对于任何离开授权团队的程序包，这些扫描发现均被视为发布阻断项。随论文附上的代码仓库仅包含明确列出的输出文件：本手稿读取的文件，以及每张图和每个表的底层文件。发布前，扫描器会检查所有列出的文件，包括统计文件；而关联分析和距离诊断的工作导出文件则由授权团队保管。图也可以通过其几何形状呈现计数。在图3的流程图中，带状图元的宽度代表学生人数，因此图中只绘制至少包含十名学生的流量，扇区大小按所绘流量确定；发布扫描还会从 PDF 中提取每个带状图元所代表的人数，并按同一最小值要求进行核查。早期的一张分析表违反了摄取规则（一张在规则明确之前建立的每名学生一行的上下文表），因此整张表被弃用，并重建为本地处理后仅发布汇总数据；这一事件也被记录下来，因为审计跟踪是研究方法的一部分。本地缓存本身仍属于受限材料：其中存放着行级数据框，且保存在代码仓库和任何发布程序包之外。生态变量在模式中明确标示其层级（C.2节）；收入在使用前会分组或排名，且从不制成表格；距离诊断仅读取邮政编码，并输出百分位数和占比。在班级规模模型中，教师标识符会在查询过程中替换为代理标识符；系统绝不读取教师姓名，也不会按教学班或教师对任何输出编制索引。

D 专业角度索引

论文正文中的弦图（图3）在每个扇区标注了对应专业的名称。表D1按字母顺序列出各专业，以及每个扇区的角位置和大小，便于在圆环上定位专业。

表D1：图3弦图的角度索引：按字母顺序列出每个专业、其扇区中心的大致角位置（从3点钟方向起逆时针度数），以及 N，即该专业在图中所示流量（至少涉及10名学生的流量）中捐出和接收的学生人数总和。名称为“通识”的专业是括号内学院面向全院的未定专业或通识学习专业：商学（COB）、工程（COE）、理学院（COS）、文理艺术学院（COLFA）、建筑、建造与规划学院（COA），以及教育与人类发展（COEHD）。“其他”汇总了总流量最小的14个专业，这些专业对应的扇区太小，无法看清：音乐；通识（COA）；人类学；非学位攻读；通识（COEHD）；地质科学；公共管理；现代语言与文学；种族、民族、性别与性研究；未定专业－GATES；COPP（行政单位）；咨询；社会工作；以及双文化／双语研究。

专业	角度	N	专业	角度	N
会计	230°	972	通识（COLFA）	119°	239
建筑学	124°	273	通识（COS）	152°	372
艺术与艺术史	96°	109	历史	105°	169
生物学	19°	3,648	信息系统与网络安全	210°	933
生物医学工程和化学工程	135°	314	跨学科学习与教学	250°	1,006
化学	185°	588	运动机能学、健康与营养学	355°	2,348
土木与环境工程	171°	503	管理学	337°	1,834
传播学	283°	1,199	管理科学与统计学	101°	160
计算机科学	295°	1,291	市场营销	308°	1,322
建筑科学	110°	176	数学	146°	350
刑事司法	193°	673	机械工程	272°	1,148
经济学	114°	213	哲学与古典学	178°	566
电气与计算机工程	201°	689	物理学与天文学	92°	85
英语	130°	279	政治学与地理学	164°	420
环境科学与生态学	141°	315	心理学	321°	1,585
金融学	261°	1,014	社会学	240°	978
普通（COB）	220°	946	大学学院	55°	5,818
普通（COE）	158°	397	其他	84°	1,396