

等级赋分制对高考考生内卷程度的影响

车天翊 储子禾 阚紫嫣 王天晓 阎芃寰 周 睿*

2023 年 3 月

摘 要：本研究聚焦于探究高考改革背景下等级赋分制对于高考考生内卷程度的影响，在参考锦标赛理论和大型竞赛理论的基础上构建理论模型，得出预测——等级赋分制不会改变主科内卷程度，只会影响选科内卷程度；一定条件下，档位密度的增加会使内卷程度有所下降；实行等级赋分制与否对内卷的影响依赖具体参数，但代入现实参数模拟的结果表明现实中实施的等级赋分制改革会降低考生总体内卷程度。随后，本研究通过 CFPS 数据库中教育支出、学业压力等变量探究高考改革对上海和浙江考生内卷程度的影响；进一步通过问卷调查获取更新、更全面的样本，覆盖更多省份，同时进行更加细致的控制变量；最后通过线上实验的方式，用打字和口算两项活动模拟高考情境，对等级赋分制的有无、疏密与内卷的关系展开了深度探究，最终验证了理论预测。

关键词：高考改革、等级赋分制、内卷

1 引言

1.1 研究背景

近年来，随着宏观经济发展速度总体放缓，“僧多粥少”的就业困难问题日益突出。为了在就业市场中取得优势地位，人们越发倾向于在教育竞争中取得优胜，这样的信念既符合千年中国科举制历史背景下的国民心理，更受到当代经济学理论的佐证——无论是教育的人力资本积累理论（Becker, 1964），还是教育的信号作用理论（Spence, 1973），都指向了人们为了就业优胜而注重教育这一结果。

因此，在中国的教育制度下，经济增速放缓、就业困难的压力集中体现在以普通高等学校招生全国统一考试（以下简称“高考”）为代表的选拔性考试中。在这样的进程中，“内卷”一词因形象地描绘了人们在激烈的教育竞争中不断追加投入的现象，在全社会急速传播，得到社会各界广泛关注。

“内卷”一词最先由人类学家 Geertz（1963）在研究印度尼西亚农业时提出，用以指代政治与科技没有发展，而社会复杂性不断提升的简单轮回状态。在经济学视角下，内卷描述的是一种在边际成本大于边际收益时仍然继续投入，直至边际收益几乎为零的状态。在经济增速放缓压力下，人们为了争取有限的教育和就业资源而激烈竞争，不断增加投入，仿佛置身“囚徒困境”之中，这恰恰与内卷的学术含义相近。在本研究中，由于各高校录取名额基本稳定，因此学生投入越多，其内卷越激烈。因而本研究主要采用投入来刻画内卷。

具体到高考考生而言，内卷这种无实质意义的消耗无疑是一件坏事——巨大的学习压力一方面可能带来心理问题，危害青少年的健康成长，另一方面也剥夺了学生们学习更多知识，拓展自己眼界的机会。为此，本研究重点关注与高考考生内卷相关的机制，期待能够在制度设计的维度上寻求增进高考考生福祉的方式。在这个过程中，等级赋分制格外引人关注。

2014 年国务院颁布了《国务院关于深化考试招生制度改革的实施意见》（国发〔2014〕35 号）（以下简称《意见》），在“启动高考综合改革试点”条目，《意见》指出“不分文理科”这一改革任务，而为了处理不区分文理科所造成的统一排名问题，等级赋分制应运而生。

*六位作者均为第一作者，名字按拼音顺序排列。每位作者的贡献相同。

等级赋分制指针对选科（除语文、数学、外语以外的科目）首先根据原始分进行排名，然后以一定的划分方式将考生划分为不同的档位，并为同一档位上的考生赋予相同的分数。以两档赋分的简单情况为例：有 10 名考生，排名前 50%（即前 5 名）归为一档，统一赋 100 分，排名后 50%（即后 5 名）的考生归为一档，统一赋 0 分，即使在每一个档位中考生的原始分数存在差异，但最终分数也是一致的。实际上等级赋分的档位越多，分差越小，赋分规则越复杂。

之所以考虑等级赋分制对高考考生内卷程度的影响，是因为朴素的直觉在思考这个问题时会发生矛盾。一方面，原本分数各异的考生会被赋分到同一等级，这可能会促使投入减少，内卷趋缓；但另一方面，赋分制可能会拉开处在档位边缘的同学之间的差距，让两个原始分相差有限的学生直接相差一整个档位，这又可能导致投入增加，内卷加剧。具体考虑赋分密度的问题——随着赋分密度的增加，一方面每一档的人数变少，考生减少努力时从本档位滑落的可能性增加；另一方面，每档分差变小，考生追求更高档位的欲望更弱，付出额外努力的可能性下降。该问题难以通过直觉得出答案，却又关乎广大处于高考改革省份的考生的福祉。因此，本研究将通过理论建模以及实证探究来对这个问题展开细致探索。

针对该问题，当前学界的相关研究极其有限。目前学界对于内卷的研究大多集中于中国农业生产、国有企业生产、政治制度领域（黄宗智，2000；王国伟，2007；王禄生，2022），而教育学领域对内卷的讨论又缺乏深刻的理论构建与细致的实证研究。而学界对等级赋分制的研究则主要侧重于其是否能满足自由选科下公平竞争的要求，缺乏对于等级赋分制其他影响的探索。

对此，本研究通过丰富的研究方法将内卷概念与等级赋分制联系起来，不但有助于揭示高考制度对考生福祉的影响，更可以拓展在各类激励机制的设计上，譬如企业绩效奖金的发放规则、赛事奖项的设置模式等等。本研究填补了学术界在此领域的空白，且具备深刻的社会价值。

1.2 研究结构

首先，本研究通过细致的政策梳理，对高考现行的等级赋分制建立了深刻的认知。通过对现有等级赋分制相关文献的总结，明确了可以深入研究的方向。

随后，本研究参考了锦标赛理论和大型竞赛理论，进行了经济学建模，给出了三点预测：第一，等级赋分制不改变主科投入；第二，满足一定条件的情况下，等级赋分制越密，考生内卷程度越低；第三，现实中的等级赋分制很可能降低考生内卷程度。

为了验证理论预测，本研究先进行了基于 CFPS 数据库的分析——聚焦上海、浙江改革前后的数据，通过教育支出、学业压力等变量刻画考生内卷程度，通过观察省份和改革的交互作用进行分析。数据展现出一定的趋势，但由于样本变异性较大等原因，这部分分析的结果并不显著。

为了获得时间更新、覆盖省份更广、样本更充足的高考考生数据，进一步明确变量间关系，本研究自主设计问卷，面向六个经历过高考改革的省份的考生发放问卷，聚焦在考生的主科时间投入和选科时间投入上展开了分析。这部分分析的结果表明高考改革后考生选科时间投入显著下降，主科投入没有显著变化，总体时间投入显著下降。这与理论预测的结果基本一致。

在问卷数据分析结果的基础上，本研究进一步将赋分制从高考改革的整体中抽离出来进行细致探索。本研究组织了线上实验，设置了模拟现实高考的激励机制，将参与者根据原始分、密集赋分和稀疏赋分三种方式分为三组，考察参与者在打字和口算项目上的投入。尽管受到样本量较小等因素的影响，实验分析的结果仍然呈现了符合理论预测的趋势——相较原始分情形，打字作为赋分科目，在赋分机制下参与者投入更低，但在赋分档位变得稀疏后投入有所回升；口算作为非赋分科目，参与者投入基本保持稳定，不受赋分机制的影响；参与者整体投入在密集赋分机制下有所下降，而在稀疏赋分机制下有所回升。这一结果增强了本研究结论的稳健性和一般性。

本文后续的部分结构安排如下：第二、三部分为政策梳理与文献综述；第四部分为理论建构；第五、六部分为目标省份选择和基于 CFPS 数据库的实证分析；第七部分为基于问卷的实证分析；第八部分为实验研究；第九部分总结了研究结论并展望了后续研究可能。

2 政策梳理

为了更好地明确研究对象与研究方法，本研究先对高考改革和等级赋分制的相关政策做了细致梳理。

2.1 高考综合改革的基本情况

高考综合改革的启动以 2014 年国务院发文《意见》为标志。实际上，《意见》中所谈及的“考试招生制度改革”，涉及范围比高考综合改革大得多；而本研究所关心的赋分制，是高考综合改革中考试科目改革的一部分，保证不同科目之间的分数具有可比性，且可以加总。整个改革的任务全貌呈现如图 2-1 所示。

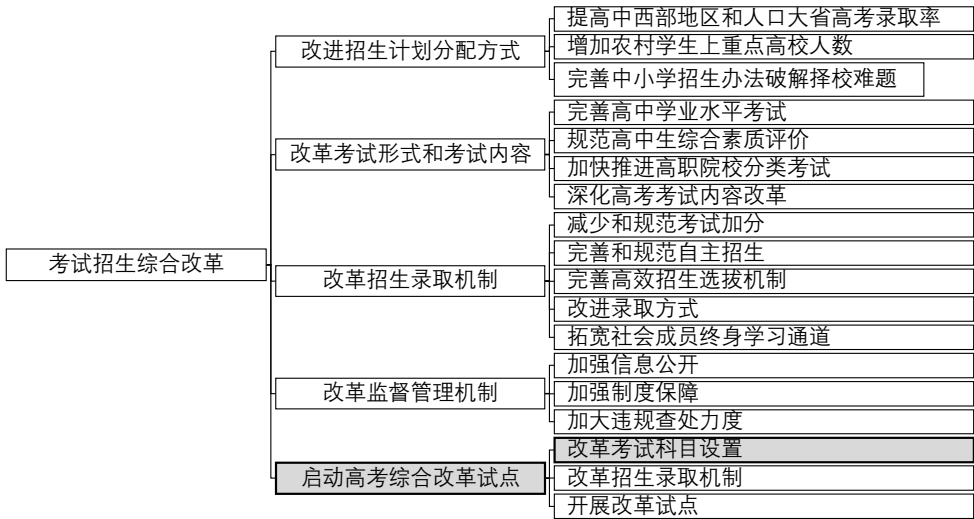


图 2-1 考试招生制度改革主要任务

高考综合改革以前的高考称为“老高考”，改革之后称为“新高考”。从“老高考”到“新高考”，对考生产生直接影响最大的一项改革内容就是考试科目改革——取消文综、理综，设立“等级性考试”科目，或称“选考”科目。

“老高考”中只有文科、理科两种选择，文科生的考试科目是语文、外语、文科数学、文综（由历史、地理、政治构成），理科生的考试科目是语文、外语、理科数学、理综（由物理、化学、生物构成）。“新高考”的考试科目则由两部分组成——“统一高考”和“等级性考试”。“统一高考”科目，即本文中所谓“主科”，包含语文、数学、外语。其中，数学不再区分文理科卷，外语逐步开始提供 2 次考试机会；“等级性考试”科目为文综、理综的替代科目，包括思想政治、历史、地理、物理、化学、生物这 6 门学业水平考试科目（浙江省还包括技术作为第 7 门科目）。考生可以根据报考高校要求和自身特长，在这些科目中自主选择 3 门作为等级性考试科目。本文将主科以外的科目统称为“选考”。

选择自由度方面，有“3+3”和“3+1+2”两种模式：“3+3”即可以从所有选考科目中任意选择 3 门，共有 20 种选择（浙江省有 35 种选择）；而“3+1+2”则要求首先从物理、历史中至少选择 1 门，然后再从剩余科目中自由选择 2 门，共有 16 种选择。

命题方面，北京、上海、天津为全自主命题，即“统一高考”和“等级性考试”所有科目试卷均由市教委自主命题。除这 3 个城市外，其它省份“统一高考”均采用国家统一命题试卷，“等级性考试”均由省教委自主命题。

2.2 赋分制规则详解

高考综合改革允许自主选科之后，自然产生了公平性问题——不同科目的内容、题型、难度、考查角度不同，考生的原始分分布不同，如果仍然采用原始分简单线性相加的算法，对于选考了难度较大、原始分分布集中于较低分数段的科目的考生，将是极不公平的。为解决公平性问题，必须采用赋分制。赋分制的基本思想是，分数直接反映考生在所有选考同一

门科目的考生中的排名，而非考生的原始作答情况。我国各省市根据自身实际情况，采用了不同的赋分制规则，大体可以分为以下三种：“同档同分”“同档内等比例转化”“标准分”。以下分别详细介绍。

2.2.1 “同档同分”或“三分一档”

“同档同分”仅针对等级性考试科目成绩进行赋分，统一高考科目仍然采用原始分。对每一门等级性考试科目成绩，按照考生原始成绩从高到低划定 A、B、C、D、E 五等，分别占各科目考试人数的相应比例；在五等级基础上进一步细化为若干级，按最接近的累计比例划定。表 2-1 和表 2-2 分别展示了两种不同的“同档同分”赋分制规则。

等	A		B			C			D		E
比例	15%		30%			30%			20%		5%
级	A+	A	B+	B	B-	C+	C	C-	D+	D	E
比例	5%	10%	10%	10%	10%	10%	10%	10%	10%	10%	5%
分数	70	67	64	61	58	55	52	49	46	43	40

表 2-1 上海市等级性考试成绩计入高考录取总成绩的等级比例和分值

等	A					B				
比例	15%					40%				
级	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5
比例	1%	2%	3%	4%	5%	7%	8%	9%	8%	8%
分数	100	97	94	91	88	85	82	79	76	73

等	C					D					E
比例	30%					14%					1%
级	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	E
比例	7%	6%	6%	6%	5%	4%	4%	3%	2%	1%	1%
分数	70	67	64	61	58	55	52	49	46	43	40

表 2-2 北京市等级性考试成绩计入高考录取总成绩的等级比例和分值

例如，假设北京市有 10,000 名考生选考物理科目，考生甲的原始分 95、市排名第 310（前 3.1%），考生乙的原始分 80、市排名第 592（前 5.92%）；考生甲、乙均位于 A3 级，所以考生甲、乙的物理成绩均为 94，尽管他们的原始分存在差异。

由于所有采用“同档同分”的省市的级间分差均为 3 分，为了直观起见，下文将“同档同分”赋分制规则统一称作“三分一档”。

2.2.2 “同档内等比例转化”或“一分一档”

“同档内等比例转化”同样仅针对等级性考试科目，统一高考科目仍然采用原始分。按照事先确定的各等级区间人数所占比例，将每门等级考试科目考生的原始成绩从高到低划分为若干等级区间；然后将每一等级区间内的考生原始成绩，依照“等比例转换法则”计算得到考生的等级成绩，公式如下：

$$\frac{Y_{max} - Y}{Y - Y_{min}} = \frac{T_{max} - X}{X - T_{min}} \quad (2.1)$$

其中 Y_{max} 、 Y_{min} 分别为该等级区间内考生原始分的最高分、最低分； T_{max} 、 T_{min} 分别为该等级区间对应分数区间的上限、下限； Y 为该考生的原始分； X 为该考生本门科目的赋分结果。

公式中，除 X 之外其它参数均为已知，可依此公式求解出 X 。此后将 X 四舍五入至整数位，作为考生该科目的最终成绩，计入高考总成绩。表 2-3 和表 2-4 分别展示了两种不同的“同档内等比例转化”赋分制规则。

等级区间	A	B+	B	C+	C	D+	D	E
比例	3%	7%	16%	24%	24%	16%	7%	3%
分数区间	[91,100]	[81,90]	[71,80]	[61,70]	[51,60]	[41,50]	[31,40]	[21,30]

表 2-3 山东省等级性考试成绩计入高考录取总成绩的等级区间和分数区间

等级区间	A	B	C	D	E
比例	15%	35%	35%	13%	2%
分数区间	[86,100]	[71,85]	[56,70]	[41,55]	[30,40]

表 2-4 河北省等级性考试成绩计入高考录取总成绩的等级区间和分数区间

例如，假设河北省有 10,000 名考生选考物理科目，考生丙原始分 80、省排名第 1,632（前 16.32%），考生丁原始分 75、省排名第 2,844（前 28.44%），则考生丙、丁均位于 B 等级区间；又假设在所有位于 B 等级区间的考生中，最高原始分为 82、最低原始分为 70，则根据计算公式， $\frac{82-80}{80-70} = \frac{85-X_{\text{丙}}}{X_{\text{丙}}-71}$ 解得， $X_{\text{丙}} = \frac{248}{3} \approx 82.67 \approx 83$ ； $\frac{82-75}{80-70} = \frac{85-X_{\text{丁}}}{X_{\text{丁}}-71}$ 解得， $X_{\text{丁}} = \frac{461}{6} \approx 76.83 \approx 77$ 。因此考生丙、丁的物理成绩分别为 83 分、77 分。

由于该赋分规则下最终成绩的分差均为 1 分，同样为直观起见，下文将“同档内等比例转化”赋分制规则统一称作“一分一档”。

2.2.3 “标准分”

“标准分”需要对全部考试科目进行分数转换，由于其转换规则复杂，且全国仅有海南省一地采用，本研究不考虑。

2.3 赋分制特点总结

相比于分数连续分布的原始分，赋分制的分数和排名结果具有两个特点。

其一，考生相对排序不变。所有赋分制规则中，等级区间和原始分排名比例都是对应的，保证了不会出现原始分更高的考生赋分排名反而较低的情况。

其二，考生绝对分数的分布改变。无论是“同档同分”还是“同档内等比例转化”，从等级区间人数比例中容易看出，这些等级并不是线性划分的，而是参考正态或偏态分布进行划分的。如果某一科目所有考生原始分的分布并非赋分制规则里所规定的那种正态或偏态分布，就将出现这样的情况：在每个等级区间内部，考生原始分之间的差距被压缩为相等或相近的赋分结果；而在两个等级区间交界处，考生原始分之间的差距则会被成倍放大。

值得注意的是，虽然原始分分数可以连续取值，但是原始分直接计入总成绩并非赋分制档位趋密后的极限情况，因为赋分制的精髓在于“赋”。一个极端例子是，满分 100 的考试里，如果按原始分看两个人分别是 60 和 61，赋分后变为 99 和 100；即使两人差距不变，赋分制对绝对分数的及其分布的改变也极有可能改变其在不同科目上投入的精力，从而影响内卷程度。

3 文献综述

为了对前述问题建构合适的理论模型，本研究参考学界现有的对于高考等级赋分制、锦标赛理论以及大型竞赛概念的研究，呈现为下面的文献综述。

3.1 高考等级赋分制的研究综述

等级赋分是对考生原始分数的一种等值处理，其遵循经典测量理论（CTT），根据考生成绩所在的等级给予考生等值的分数。事实上，不少国家及地区的升学考试都在采用等级赋分制度，比如英国的普通中等教育证书高级水平考试（A-Level）、德国的高中毕业考试（Abitur）、法国的高中毕业会考（BAC）等。就我国而言，2014 年《意见》正式颁布，其对高考科目设置作出了重大制度创新，选考科目赋分制度改革随之跟进。通过对原始分数进行相应的处理，实现不同科目的分数可直接加总得到一个相对可比的分数，是等级赋分方式设计的初衷。而从国家对新高考改革的总体设计与规划来看，通过等级赋分来弱化微小分差对考生考试结果的决定性影响是未来发展趋势，符合新时代深化教育评价改革的总体要求。

等级赋分的设计主要遵循科学性和公平性两个原则，学者们对此展开了很多讨论。程伟等（2020）比较分析了三种赋分方式的区分程度，认为不同省市应制定适合自身情况的赋分方式；陈逸清（2021）从公共管理的角度分析总结了我国各阶段高考招生制度中的公平性问题，重点剖析归纳上海新高考改革的可取之处、存在的问题和原因；罗立祝（2020）详细阐明两种等级赋分方案各自的优劣，并对完善现有选考科目赋分机制提出了新的方案。实践的开展推动新的研究的产生，在观察到浙江高考物理考生人数锐减的状况后，于涵等（2018）提出赋分使考生的科目选择不可避免地产生驱赶效应和磁吸效应，而其他学者则进一步聚焦赋分机制的设计：温中麟（2017）建议使用校准分数假设关系方案，北京教育学会倡议基于“百分位等级”的分段转换方案（臧铁军和杨君，2017），李金波（2019）提出将固定占比的等级赋分转变为浮动占比的等级赋分方案，等等。

总的来看，高考等级赋分制度是一个较新的话题，而当前该领域的研究主要聚焦探讨赋分是否科学地反映了考生的实力、是否实现了教育公平等问题（黄海龙，2020；郑若玲等，2020），主要聚焦在技术细节上，对考生自身的反应和行为决策分析较少，更没有研究将等级赋分制与内卷联系起来。本研究则会着重关注考生面对不同计分方式进行的行为决策，站在整个社会的角度，思考高考改革的意义与影响。

3.2 锦标赛理论的研究综述

本研究在理论建模部分参考了锦标赛理论模型。

锦标赛是参与者争夺根据相对排名颁发的奖品的一种竞赛。锦标赛理论源于体育竞技中的锦标赛，是通过比较相对业绩来研究委托代理关系的一种激励机制理论。公司中员工报酬水平随职位提升而呈现出阶梯式跃动的现象是如人力资本理论、职业生涯激励理论等传统理论无法解释的，因而 Lazear 和 Rosen 于 1981 年共同提出了新的思考框架。

在该模型中，两个风险中性的代理人构成竞争，雇主根据代理人业绩的高低对他们进行排名，以此确定各自的报酬。锦标赛理论的核心特征是关注代理人的相对排名而非绝对绩效，以此调动代理人自发竞争，从而形成激励。其基本结论是，代理人投入的努力程度主要受业绩排名竞赛中获胜者和未获胜者的薪酬之差以及外界环境不确定性因素的驱使，两个因素共同决定了最终绩效成果的产出。由于职位越高，获得的报酬就越丰厚，所以员工会为了不断提高的职位报酬而努力工作。正是由于这种薪酬溢价的存在，加大 CEO 同其他高层管理成员之间的薪酬数额的差距将会降低委托人对代理人的监控成本，给委托人和代理人之间的利益一致性提供强激励，促使有才能的高管付出努力向上晋升，创造相对更好的公司绩效。这就好像锦标赛中运动员为了高额奖金而不断努力提高个人排名的过程。

继该基础模型提出以后，锦标赛理论迅速发展。学者对锦标赛理论开始了不同方向的研究。首先，一些学者将锦标理论与计件工资理论、合同理论等一些传统理论进行了比较分析，进一步从理论上证明了锦标赛的优越性（Green & Stokey, 1983; Budde, 1984; Kräkel, 2005; Tsoulouhas & Marinakis, 2007）；其次，某些文献探讨了锦标制度的设计问题，主要侧重于研究报酬水平和报酬差距对绩效产出的影响等（O'keeffe et al., 1984; Fehr & Falk, 2002; Harbring & Irlenbusch, 2003; Orrison et al., 2004）。这些研究得出一个核心结论：提高报酬和增大薪酬差距有利于提升代理人的努力水平。同时，锦标赛模式也从个体层面逐渐扩展到组织团队层面，团队锦标机制渐渐成为研究的热点（Heneman & Von, 1995; Dickinson & Isaac, 1998; Gneezy & Rustichini, 2000; Irlenbusch & Ruchala, 2008）。

相比国外而言，锦标赛理论在国内的研究起步较晚。代逸生和林科（2013）是国内较早系统结合锦标赛理论进行行业研究的学者之一，两人应用数理模型对锦标赛理论的来龙去脉进行了推理剖析。近些年来，锦标赛也被应用到中国地方政府的晋升模式上，并对改革开放 30 年来持续快速的经济增长作出有力的解释（周黎安等，2005；周黎安，2007）。

经过三十多年的发展,锦标赛理论的很多结论已被验证。锦标赛理论的现有研究主要是从锦标赛的有效性、怎样克服锦标机制中的潜在问题和如何设计锦标模式等方面展开(杨华, 2019; 余静, 2020; 马秋敏, 2019; 魏光兴和陈永恒, 2022)。这些研究对体育锦标赛、企业员工激励以及政府官员的晋升竞争都具有重要的指导意义,也能为我国激励制度的改革提供新的思路。特别地,在锦标赛理论中,奖励结构基于相对排名而非绝对产出水平这一核心思想极为契合高考的情境,因而本研究在进行理论建构时着重参考了相关理论。

3.3 大型竞赛的研究综述

大型竞赛的模型在求解大量竞争者行为方面作出了突出贡献,为本研究的模型求解提供了重要参考。

大型竞赛的概念由博弈论学者 Olszewski 与 Siegel (2016) 提出,模型注重分析有大量异质的参与者进行竞争的情况。该模型的建立使原本复杂的精确均衡表征或没有现存均衡表征的竞赛均衡得以拥有近似解,扩展了对具有不对称竞争者和异质激励的比赛的认知和理解,也对竞赛设计、福利分析和比较静态都有很大贡献。

Olszewski 和 Siegel 持续对该模型进行完善和拓展,如探讨了不同类型的“投标上限”对大型竞赛中所有参与者的加总成本的影响(2019);研究了促使参与者绩效最大化的激励设置结构(2020),不仅对现实世界竞赛的奖品设置提出改进意见,也为制定和解决迄今为止已证明难以解决的其他竞赛设计问题提供新的思路。

大型竞赛对大量选手参与的竞争与博弈进行分析,对高考这种大量考生共同竞争的考试有很好的借鉴意义。已有研究从大型竞赛或帕累托改进的视角分析高考录取的话题,提出考生竞争大学录取名额的福利成本损失的问题(Bodoh & Hickman, 2018; Krishna et al., 2022)。因此相关理论为本研究进行理论建模提供了指引与参考。

3.4 文献综述总结

就本研究的贡献而言,本研究拓展了锦标赛理论和大型竞赛的模型在我国高考制度上的应用,构建起针对性的理论模型,并在全国范围内开展实证研究,以理论和实证相结合的方式对社会热点问题展开深入思考,为日后的高考改革乃至赋分制在更大范围内的应用和拓展提供了研究基础。

4 理论分析

4.1 概念界定

如前所述,高校录取的名额仅按最终考试的相对排名发放,不随各届考生的绝对分数变化;可以认为无论考生在备考时投入程度如何,全体考生的总收益保持不变。面对社会升学机会总数不变的情况,考生相互裹挟,不断增加投入(包含学习时间、练习题量、家庭支持等各种形式),这符合内卷的定义。因此,高考考生内卷程度可以由全体考生的总投入度量,考生总投入越多,内卷程度越高。

本研究的理论模型、行为实验与实证分析均以考生备考总投入(或平均投入)作为内卷程度的主要衡量标准。

4.2 模型建构

等级赋分制对高考考生内卷程度的影响可以由锦标赛理论(Lazear & Rosen, 1981)部分推断。锦标赛理论认为,在一个每个人选择自己的表现水平、最终根据表现水平在人群中的排名发放相应奖励的锦标赛中,相邻排名之间的奖励差距越大,人们选择的表现水平就越高,投入的精力也越多。这意味着在两个档位之间的考生的内卷程度会随着档位密度的增大而减小(因为档位之间等级分差距减小)。但传统的锦标赛理论无法预测等级赋分制档位密度对所有考生内卷程度的综合影响。完整呈现各种等级赋分制的影响,需要用到大型竞赛理论。

本文基于 Krishna 等(2022)的研究建立模型。简单起见,假设高考中仅有一门主科和一门选科,其中主科得分始终为原始分,选科得分为原始分或等级分。每一位考生选择自己在主科和选科分别希望达到的目标原始分 t_M 与 t_m , 并付出相应的精力投入 $c_M(t_M)$ 和 $c_m(t_m)$,

其中 c_M 和 c_m 均二阶可导, 且有 $c_M(0) = c_m(0) = 0$ 和 $c'_M(x), c''_M(x), c'_m(x), c''_m(x) \geq 0$ 。考生的高考总分为 $t_M + \phi(t_m)$, 其中函数 ϕ 将选科的原始分转换为等级分。当选科采用原始分时, $\phi(t) = t$ 。

为了刻画考生能力禀赋的异质性, 假设每位考生分别属于某种类型 $x, x \in [0, 1]$ 。用 $g_1(x)$ 和 $g_2(x)$ 两个函数刻画考生高考得分的边际效用以及考生的学习能力, 满足 $g_1(x), g_2(x) \geq 0$ 。类型为 x 的考生高考每高一分, 其效用增加 $g_1(x)$ 单位。类型为 x 的考生想要在某一科目达到 t 分的成绩, 需要的投入为 $\frac{c(t)}{g_2(x)}$ 。可以对 x 进行排序使得 $g_1(x) \cdot g_2(x) = x$ 恒成立。假设 $g_1(x)$ 关于 x 单调递增, 即 $g'_1(x) \geq 0$ 。

记 x 的累积分布函数为 $F(x)$, 密度函数为 $f(x)$, 则 $f(x) = F'(x)$ 。假设 x 的分布是对称分布, 即 $f(x)$ 关于 $x = \frac{1}{2}$ 对称。假设 $f(x)$ 在 $[0, \frac{1}{2}]$ 上关于 x 单调递增, 即 $f'(x) \geq 0, \forall x \in [0, \frac{1}{2}]$ 。这意味着考生的能力禀赋分布是一个中等水平较多、极好和极差水平较少的类似正态分布的分布, 符合现实直观。

在上述假设下, 考生的效用函数形式如下:

$$U_{total}(t_M, t_m; x) = g_1(x) \cdot (t_M + \phi(t_m)) - \frac{c_M(t_M) + c_m(t_m)}{g_2(x)} \quad (4.1)$$

考生总投入为:

$$C_{total} = \int_0^1 \frac{c_M(t_M) + c_m(t_m)}{g_2(x)} f(x) dx \quad (4.2)$$

4.3 模型求解

4.3.1 主科

命题 1 最大化 U_{total} 的 t_M 仅与 x 和 c_M 有关。

命题 1 证明:

记 $U_M(t_M; x) = g_1(x)t_M - \frac{c_M(t_M)}{g_2(x)}$, $U_m(t_m; x) = g_1(x)\phi(t_m) - \frac{c_m(t_m)}{g_2(x)}$, 则有:

$$U_{total} = U_M + U_m \quad (4.3)$$

最大化考生效用的 t_M 满足一阶条件:

$$\begin{aligned} 0 &= \frac{\partial U_{total}}{\partial t_M} \\ &= \frac{\partial U_M}{\partial t_M} + \frac{\partial U_m}{\partial t_M} \\ &= \left(g_1(x) - \frac{c'_M(t_M)}{g_2(x)} \right) + 0 \end{aligned} \quad (4.4)$$

求解一阶条件得到 $t_M = c'^{-1}_M(x)$, 仅与 x 和 c_M 有关。证毕。

命题 1 表明考生的效用可以拆分为主科和选科两个独立的部分, 考生在两个部分内分别最优化自己的选择。

类似地, 考生总投入 C_{total} 也可以拆分成主科和选科两部分:

$$\begin{aligned} C_{total} &= \int_0^1 \frac{c_M(t_M) + c_m(t_m)}{g_2(x)} f(x) dx \\ &= \int_0^1 \frac{c_M(t_M)}{g_2(x)} f(x) dx + \int_0^1 \frac{c_m(t_m)}{g_2(x)} f(x) dx \end{aligned} \quad (4.5)$$

记 $C_M = \int_0^1 \frac{c_M(t_M)}{g_2(x)} f(x) dx$, $C_m = \int_0^1 \frac{c_m(t_m)}{g_2(x)} f(x) dx$, 则有:

$$C_{total} = C_M + C_m \quad (4.6)$$

在 f 和 c_M 确定的情况下, C_M 为常数。这说明选科的赋分制度对考生内卷程度的影响完全体现在考生在选科的投入上, 与主科无关。简便起见, 以下分析中以 U_m 作为考生的效用函数, 以 C_m 度量考生总投入 (即内卷程度), 并省略所有角标上的 m 。

4.3.2 等级赋分制

在等级赋分制下, 选科的最终得分 $\phi(t)$ 可以表示为 $\phi(t, \mathbf{t}_-)$, 其中 \mathbf{t}_- 为其他所有考生选择的目标原始分组成的向量。此时, 考生的效用函数为:

$$U(t, \mathbf{t}_-; x) = g_1(x)\phi(t, \mathbf{t}_-) - \frac{c(t)}{g_2(t)} \quad (4.7)$$

这样的多人博弈模型难以直接求解纳什均衡, 但可以利用 Olszewski 和 Siegel (2016) 对大型竞赛的研究结论, 用一个直接按照类型 x 从高到低分配分数的单一主体机制对上述多人非对称博弈进行近似。具体而言, Olszewski 和 Siegel 证明了类型 x 的博弈主体会以极大的概率选择如下目标分数:

$$t(x) = c^{-1} \left(x \cdot y(x) - \int_0^x y(\tilde{x}) d\tilde{x} \right) \quad (4.8)$$

其中函数 $y(x)$ 表示直接按照考生类型从高到低进行等级赋分, 而非通过对考生选择的目标原始分进行排序后再赋分。记 $G(y)$ 为等级分的累积分布函数, 则 $y(x)$ 为以下关系式确定的隐函数:

$$G(y) = F(x) \quad (4.9)$$

该关系式意为考生能力禀赋在人群中的百分位等于赋给该考生的分数在全部档位中的百分位。记 $G^{-1}(z) = \inf\{y | G(y) \geq z\}, z \in [0, 1]$, 则 $y(x) = G^{-1}(F(x))$ 。简单起见, 以下讨论的等级赋分制中, 相邻档位的等级分差均相等, 且每个档位人数均相等。

考虑一个满分为 β , 共 n 档 ($n \in \mathbb{Z}, n \geq 2$) 的等级赋分制。记均衡时类型 x 考生选择的目标原始分为 $t_n(x)$, 等级分的累积分布函数为 G_n , $y_n(x) = G_n^{-1}(F(x))$, 则有:

$$G_n^{-1}(z) = \beta \frac{i}{n-1}, \quad \forall i \in \{0, 1, \dots, n-1\}, z \in \left[\frac{i}{n}, \frac{i+1}{n} \right) \quad (4.10)$$

规定 $G_n^{-1}(1) = \beta$ 。此时, 考生在选科上的总投入 C_n 为:

$$\begin{aligned} C_n &= \int_0^1 \frac{c(t_n(x))}{g_2(x)} f(x) dx \\ &= \int_0^1 g_1(x) y_n(x) f(x) dx - \int_0^1 \frac{f(x)}{g_2(x)} dx \int_0^x y_n(\tilde{x}) d\tilde{x} \end{aligned} \quad (4.11)$$

命题 2 C_n 关于 n 单调递减。

命题 2 证明见附录 A。

命题 2 表明在等级赋分制中, 赋分科目满分不变的情况下, 随着档位密度的增加, 全体考生在选科上的总投入下降, 内卷程度下降。

4.3.3 不赋分

选科不采用等级赋分制时, 考生的原始分就是其最终得分, 即 $\phi(t) = t$ 。此时, 考生的效用函数为:

$$U(t; x) = g_1(x)t - \frac{c(t)}{g_2(x)} \quad (4.12)$$

最大化考生效用的 t 满足一阶条件:

$$0 = \frac{\partial U}{\partial t} = g_1(x) - \frac{c'(t)}{g_2(x)} \quad (4.13)$$

求解一阶条件得到对于类型 x 的考生，最大化其效用的目标原始分 t_0 满足：

$$t_0(x) = c'^{-1}(x) \quad (4.14)$$

此时，全体考生的总投入 C_0 满足：

$$C_0 = \int_0^1 \frac{c(t_0(x))}{g_2(x)} f(x) dx = \int_0^1 c(c'^{-1}(x)) \frac{f(x)}{g_2(x)} dx \quad (4.15)$$

随着 β 、 n 以及 $c(t)$ 形式的变动，采用等级赋分制和不采用等级赋分制时全体考生的总投入的大小关系并不固定，且在积分形式下难以比较。下面为 β 、 n 、 c 和 f 等赋予具体数值或函数形式，通过数值计算的方法给出该模型对是否采用等级赋分制对内卷程度的影响的预测。

$f(x)$	n	β	C_0 不赋分内卷程度	C_n 赋分制内卷程度	$\left(1 - \frac{C_n}{C_0}\right) \cdot 100\%$ 赋分制改革减少的 考生总投入百分比 (负数为增加)
f_U	20	0.6	0.2500	0.1571	37.16%
		0.7		0.1833	26.68%
		0.8		0.2095	16.21%
		1		0.2618	-4.74%
	70	0.6		0.1521	39.16%
		0.7		0.1774	29.02%
		0.8		0.2028	18.88%
		1		0.2535	-1.40%
$f_{N, \frac{1}{2}}$	20	0.6		0.1630	34.82%
		0.7		0.1901	23.96%
		0.8		0.2173	13.09%
		1		0.2716	-8.63%
	70	0.6		0.1578	36.89%
		0.7		0.1841	26.37%
		0.8		0.2104	15.85%
		1		0.2630	-5.18%
$f_{N, \frac{1}{6}}$	20	0.6		0.2157	13.71%
		0.7		0.2517	-0.67%
		0.8		0.2876	-15.05%
		1		0.3595	-43.82%
	70	0.6		0.2112	15.53%
		0.7		0.2464	1.45%
		0.8		0.2816	-12.63%
		1		0.3520	-40.79%

表 4-1 各种参数组合下是否赋分对内卷程度的影响

为 β 赋予的值有 0.6、0.7、0.8 和 1。其中 $\beta = 0.6$ 对应了北京、浙江等省份的实际政策， $\beta = 0.7$ 对应了福建、广东等大部分第三批新高考改革省份的实际政策， $\beta = 0.8$ 对应

了山东省的实际政策^①。 $\beta = 1$ 对应了赋分前后分数变化区间不变的简单情况。

为 n 赋予的值有 20 和 70。 $n = 20$ 对应了大多数三分一档改革省份的实际政策， $n = 70$ 对应了大多数一分一档改革省份的实际政策。

为 $c(t)$ 赋予的形式为 $c(t) = c_k(t) = \frac{1}{2}t^2$ 。该形式满足前述模型中对 c 的假设，并且满足 $c'(0) = 0$ 和 $c'(1) = 1$ ，即类型 0 的考生会选择目标原始分 0，类型 1 的考生会选择目标原始分 1。可以认为这控制了不赋分时，选科的满分为 1。同时，该形式还满足 $c(1) = \frac{1}{2}$ 。对于等级赋分制中得到 β 分满分的同学，其在选科上的投入为 $\frac{\beta}{2}$ ，数值上为满分的一半，即 $\forall x \in \{x|y_n(x) = \beta\}, c(t(x)) = \frac{\beta}{2}$ 。 $c(1) = \frac{1}{2}$ 保证了无论是否赋分，想要取得满分所需的投入水平不变。

为 $f(x)$ 赋予的形式为 $f(x) = f_U(x) = 1$ 或 $f(x) = f_{N,\sigma}(x) = z_\sigma(x) + 2Z_\sigma(0)$ ，其中 $z_\sigma(x)$ 和 $Z_\sigma(x)$ 分别为正态分布 $N(\frac{1}{2}, \sigma^2)$ 的概率密度函数和累积分布函数。 f_U 的形式为一个均匀分布，而 $f_{N,\sigma}$ 的形式为一个 $[0, 1]$ 上截断双尾的正态分布。除均匀分布 f_U 外，在上述截断双尾的正态分布形式下分别取 $\sigma = \frac{1}{2}$ 和 $\sigma = \frac{1}{6}$ 进行实际计算，分别对应从一倍标准差处截断和从三倍标准差处截断，对应的 f 分别为 $f_{N,\frac{1}{2}}$ 和 $f_{N,\frac{1}{6}}$ 。

取 $g_1(x) = 1$ ， $g_2(x) = x$ 。这表明不同能力禀赋的考生对高考分数的效用评价一致，能力禀赋只影响其达到目标分数所需的投入多少，符合直观。

代入上述共 24 组不同的参数组合进行数值计算得到表 4-1。

表 4-1 表明，在大多数参数组合下，采用等级赋分制相比不赋分可以降低考生总投入，降低内卷程度。例如 $(n, \beta) = (20, 0.6)$ 对应北京改革后的政策，模型预测北京在改革为等级赋分制后，在假设了考生能力分布的情况下，考生内卷程度将降低约 13.71%。

4.4 理论预测

综合以上分析，理论模型对等级赋分制对高考考生内卷程度的影响有如下预测：

- 1) 等级赋分制仅改变考生在采用等级赋分制的科目上的投入，不改变考生在其他科目上的投入；
- 2) 在等级赋分制满分一定的情况下，赋分档位密度越大，高考考生的内卷程度越小；
- 3) 现实中实施的等级赋分制改革大概率有助于降低高考考生的内卷程度。

5 实证分析的主要目标省份选择

为了对理论预测进行验证，进行实证研究是必要的，为此，本研究先对各省级行政区的高考政策展开梳理，选取符合要求的目标省份。

2014 年起，各省、直辖市、自治区政府以《意见》中高考综合改革部分的内容为根据，结合地方实际情况，开始分批出台高考综合改革方案文件。目前为止共有 5 批高考综合改革省份，覆盖了除西藏、新疆、港澳台之外全国所有其它省级行政区。其中第 4 批、第 5 批分别至 2024、2025 年起实施，无法纳入到研究范围内。

前 3 批改革省份的具体的改革时间和改革措施如表 5-1 所示。上海、北京、天津自改革以来一直采用“同档同分”；山东和所有第 3 批改革省份全部采用“同档内等比例转化”。海南和浙江的情况较为特殊。海南省是全国唯一采用标准分、对所有考试科目分数均进行标准化的省份，分数转换过程较为复杂，不太适合作为目标省份；而浙江省进行了二次改革，自 2022 年起从“三分一档”变成了“一分一档”。

考虑到以上情况以及小组成员可接触到的样本地域范围，本研究最终选取了“三分一档”的所有省份（上海、北京、天津）、“一分一档”的部分省份（山东、福建）和经过二次改革的浙江，共计 6 个省及直辖市，作为实证研究的主要目标省份。同时兼顾其它省份的数据。

^① 各省高考改革后等级赋分科目名义满分大多为 100 分，但其最低档分数并非 0 分。例如北京高考中赋分科目满分为 100，最低档分数为 40，其等效满分只有 $100 - 40 = 60$ 分，占原始分满分 100 分的 0.6。详见表 5-1。

改革批次	改革时间		省份	等级性考试选科模式	等级性考试赋分制度				
	启动年份	实施年份			起止分数	等级区间数量	等级区间分差	区间内分数转化	最终有效分差
1	2014	2017	上海 浙江	3+3 3+3	[40,70] [40,100]	5 等 11 级 5 等 21 级	3 3	否 否	3 3
(1.5)	2020	2022	浙江	3+3	[40,100]	5 等 20 级	3	等比例转化	1
2	2017	2020	北京	3+3	[40,100]	5 等 21 级	3	否	3
			天津	3+3	[40,100]	5 等 21 级	3	否	3
			海南	3+3	[60,300]	不划分区间	—	标准分	1
			山东	3+3	[20,100]	8	10	等比例转化	1
3	2018	2021	河北	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1
			辽宁	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1
			江苏	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1
			福建	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1
			湖北	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1
			湖南	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1
			广东	3+1+2	[30,100]	5	17,11,11,17,10	等比例转化	1
			重庆	3+1+2	[30,100]	5	14,14,14,14,10	等比例转化	1

表 5-1 前 3 批高考综合改革省份部分改革方案

6 基于 CFPS 数据库的实证分析

在完成理论建构之后，本研究首先充分利用 CFPS 数据库对理论部分的预测进行初步的检验和探索。

6.1 CFPS 数据库的介绍

中国家庭追踪调查（China Family Panel Studies, CFPS）是北京大学中国社会科学调查中心的三个大型社会调查项目之一。调查的家庭来自于科学严谨的抽样选择且不能随意替换，保证了样本的科学性和代表性；这些家庭中的所有成员，及其今后生养或领养的子女被称作“基因成员”，他们都会被纳入项目调查的范围，是项目永久追踪的对象。在中国，家庭扮演着代际关系的最重要桥梁，是经济活动与社会交往的基本单位。在关注家庭之外，项目还收集了个体和社区层面的数据。如此一来，调查所产生的数据就可以如实反映中国社会、经济、人口、教育和健康的变迁，为学术研究和公共政策分析提供基础。CFPS 数据库的特点如表 6-1 所示。

特点	描述
数据来源	有所侧重地收集了中国 25 个省、自治区、直辖市的信息，并覆盖城市和农村地区。调查对象包含样本家户中的全部家庭成员。
数据类型	涵盖了个人、家庭、社区，教育、健康、经济等多个领域的的数据。
数据时间跨度	始于 2010 年，每两年开展一次追踪调查，最新至 2020 年。
数据量	2010 年最初选定 15000 户目标样本，2022 年时已增至 22585 户。目前收集了超过 20 万个观测。
数据质量	采用了多层次随机抽样设计，具有较好的代表性。同时，CFPS 数据库还对问卷回收的数据进行了专业的清洗和加工，提高了数据的准确性和可靠性。

表 6-1 CFPS 数据库的特点

总之，CFPS 数据库是一个重要的社会科学研究资源，其多维度、大样本、高质量的特点，为研究中国社会和经济的发展提供了重要的数据支持。

6.2 数据分析方法

CFPS 现阶段便于获取的数据截止到 2018 年，从上文对各省赋分制度的政策梳理可知，六个目标省份中只有上海市和浙江省在 2018 年之前进行了高考改革。因此，基于 CFPS 的实证研究聚焦在 2017 年上海和浙江开始实行的高考改革，将其作为一次准自然实验，分析其对于高中生内卷程度的影响。因此，研究将以年份和省份作为分组变量进行二因素独立测量方差分析（Two-Way Independent-Measures ANOVA）^①，重点观察年份和省份的交互作用是否显著。若显著，则说明高考改革对内卷指标存在影响。

6.3 样本情况

本研究以年份和省份为分组变量，对于能代表学生的内卷程度的因变量展开二因素方差分析，重点观察年份和省份的交互作用是否显著。对于年份，重点关注 2016 年和 2018 年，它们横跨了 2017 年上海和浙江的高考改革。对于省份，进行如图 6-1 的划分。

对于 2016 年和 2018 年的 CFPS 数据，筛选出其中的高中生^②，得到全国范围内的样本分布如图 6-2 所示。

按照分组变量的划分，样本量分布如图 6-3 所示。

综合看出，总体上数据的样本量是符合分析要求的；但 CFPS 在各地区的样本大致均匀分配，因此上海和浙江的样本量较小，这提示着对于部分结论也要保持谨慎态度。

^① 理论上，进行 DID 分析会是最理想的，因为 CFPS 的数据是面板数据。但每个人的成长是不可逆的，有且仅有经历一次高考，每年参加高考的并非同一批考生。因此我们转而采用混合横截面数据的处理方式。

^② CFPS 数据库收集了受访者的目前最高学历和目前求学或离校阶段。当受访者目前最高学历为初中，目前求学或离校阶段为高中时，说明受访者是高中生。

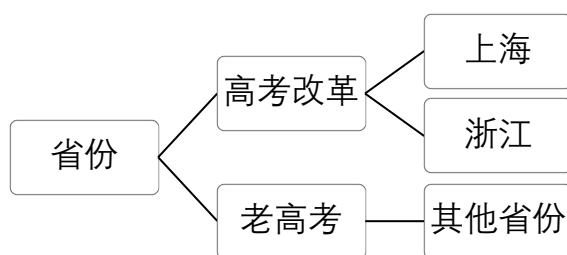


图 6-1 CFPS 数据库实证研究对省份的划分

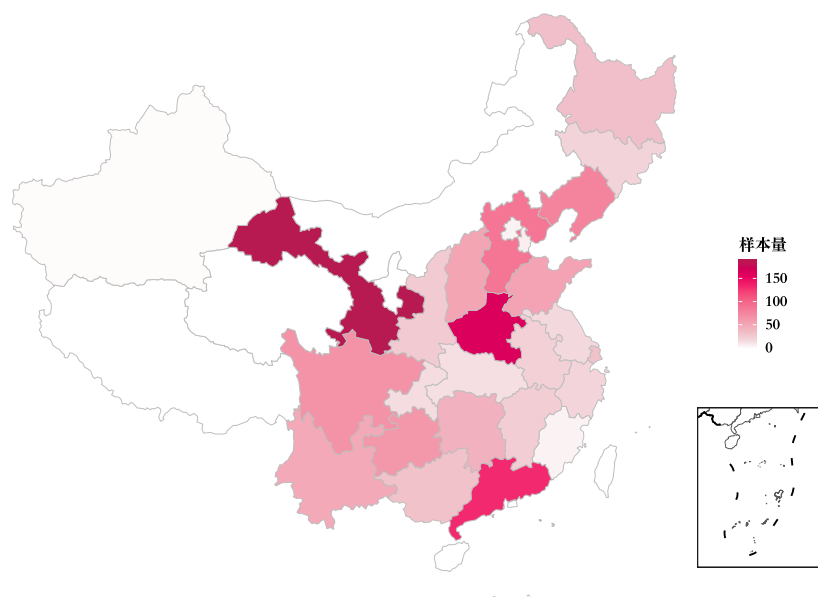


图 6-2 2016 和 2018 年 CFPS 数据库全国高中生样本来源分布

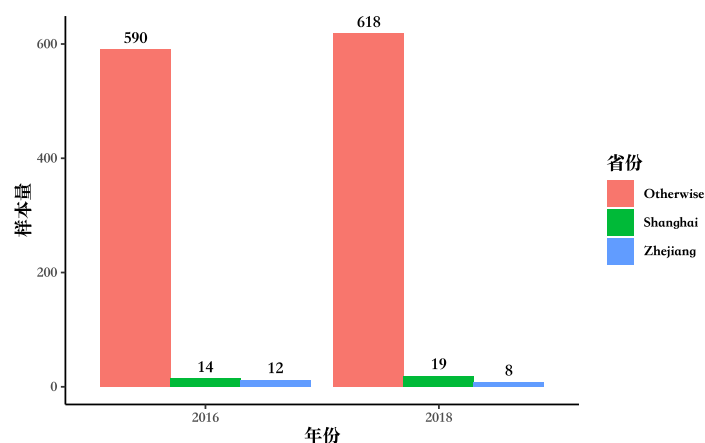


图 6-3 2016 和 2018 年 CFPS 数据库高中生样本来源分布（按组别）

6.4 变量确定

对于学生的内卷程度，相对应地选择了一系列与之相关的指标，大致可以分为刻画投入的直接指标和刻画情绪的间接指标。直接指标包括教育支出、睡眠时长、辅导班参加情况等，间接指标包括学业压力、愉悦指数、情绪低落等。这些变量在 CFPS 数据集中的名称

和定义如表 6-2 所示。

CFPS 变量名	年份	调查对象	对应问卷问题
<i>ks90total</i>	2010	成人	您去年的教育支出合计（元）
<i>wd5total</i>	2010	孩子	去年全年，这个孩子的所有教育支出总计（元）
<i>wd5total_m</i>	2012	孩子	过去一年教育总支出（元）
<i>ks9total</i>	2014	成人	过去 12 个月教育总支出（元） （系统自动加总）
<i>wd5total</i>	2014	孩子	过去 12 个月教育总支出（元）
<i>pd5total</i>	2016	成人	去年教育支出总额（元）
<i>pd5total</i>	2016	孩子	过去 12 个月教育总支出（元）
<i>pd5total</i>	2018	成人	教育支出总费用（元）
<i>wd5total</i>	2018	孩子	过去 12 个月教育总支出（元）
<i>ks502</i>	2010 至 2016	成人、孩子	学业压力
<i>qs502</i>	2018	成人、孩子	学业压力
<i>pn406</i>	2018	成人、孩子	我感到情绪低落
<i>qn406</i>	2016	成人、孩子	我感到情绪低落
<i>pn412</i>	2018	成人、孩子	我感到愉快
<i>qn412</i>	2016	成人、孩子	我感到愉快

表 6-2 CFPS 数据集变量名称及定义

对 CFPS 变量进行归类、处理，得到后文分析使用的变量名称及定义如表 6-3 所示。

变量类型	变量名称	符号	定义
被解释变量	教育支出	<i>eduexp</i>	受访者报告的过去一年家庭教育支出（元）。
	对数教育支出	<i>lneduexp</i>	教育支出的自然对数，即 $\ln(eduexp)$ 。
	学业压力	<i>pressure</i>	受访者报告的学业压力（范围：1-5）。
	情绪低落	<i>depression</i>	受访者报告的情绪低落程度（范围：1-4）。
	生活愉快	<i>happiness</i>	受访者报告的生活愉快程度（范围：1-4）。
解释变量	年份	<i>year</i>	受访者接受调查的年份。
	省份	<i>province</i>	受访者所在省份。对于上海的受访者，取值为 Shanghai；对于浙江的受访者，取值为 Zhejiang；对于其他受访者，取值为 Otherwise。
辅助变量	家庭纯收入	<i>faminc</i>	受访者报告的过去一年家庭纯收入（元）。
	家庭人均纯收入	<i>faminc_{per}</i>	受访者报告的过去一年家庭人均纯收入（元）。
	家庭总储蓄	<i>famsav</i>	受访者报告的家庭总储蓄（元）。

表 6-3 CFPS 数据分析使用的变量及其定义

正文部分将会重点展现以教育支出为代表的直接指标和以学业压力代表的间接指标的 ANOVA 结果，其他指标和处理的结果见附录 B。

6.5 数据分析结果

6.5.1 教育支出

对于教育支出，值得关注的应该是教育支出的相对增长率，而不是绝对增长量。因此，对教育支出取对数（记为 $\ln eduexp$ ）^①，处理后的分布接近正态分布，如图 6-4 所示。

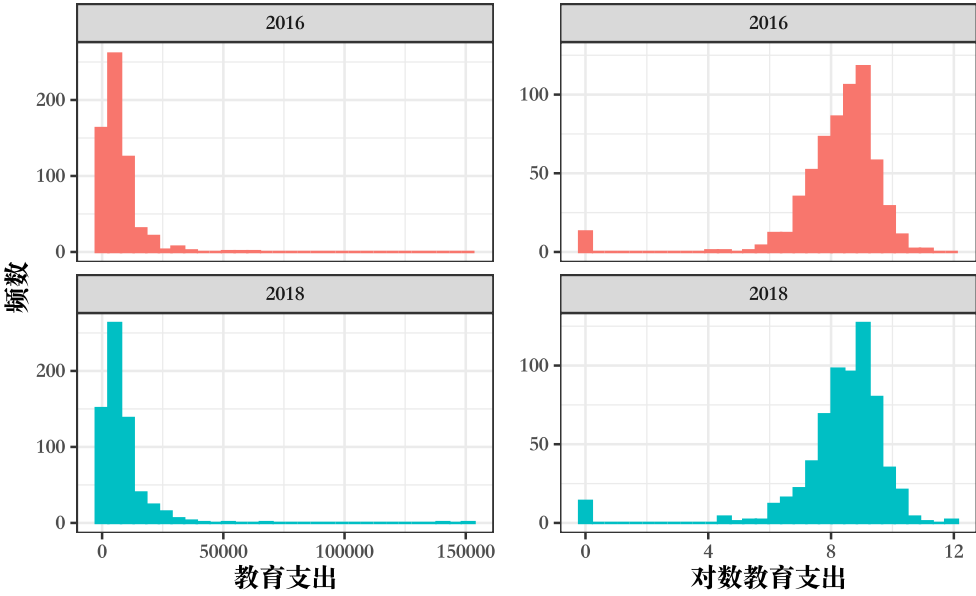


图 6-4 教育支出分布

取对数的教育支出（ $\ln eduexp$ ）的 ANOVA 结果如表 6-4 所示。

ANOVA Table							
Dependent Variable: $\ln eduexp$							
Between-Subjects Factor: $year, province$							
	MS	$MSE^{②}$	df_1	df_2	F	p	η^2
$year * province$	4.287	2.486	2	1255	1.724	0.179	0.003
Levene's Test for Homogeneity of Variance							
	F		df_1	df_2		p	
DV: $\ln eduexp$	1.477		5	1255		0.194	
Descriptive Statistics							
$year$	$province$	M	SD	n			
2016	Shanghai	8.643	2.604	14			
2016	Zhejiang	7.740	2.743	12			
2016	Otherwise	8.239	1.485	590			
2018	Shanghai	9.097	1.531	19			
2018	Zhejiang	9.092	1.124	8			
2018	Otherwise	8.323	1.610	618			

表 6-4 取对数后教育支出的二因素方差分析

① 经检验，存在 41 个样本的教育支出为 0，这些样本取对数后赋值为 0。

② MSE 指的是均方误差（Mean Squared Error）。后同。

从图像上看，2016 年和 2018 年各地区的取对数的教育支出 ($\ln eduexp$) 的变化趋势如图 6-5 所示。

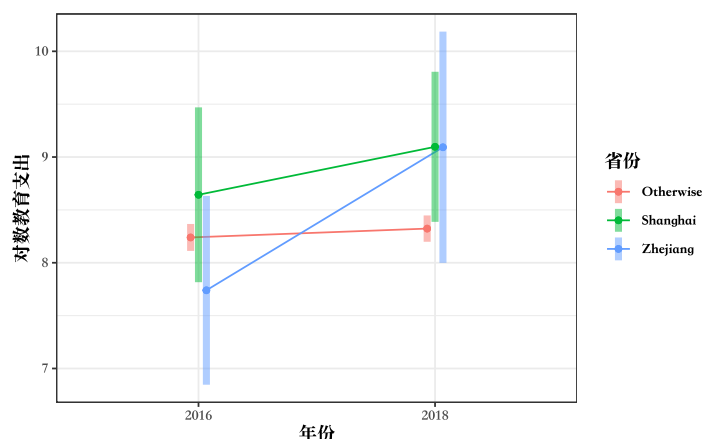


图 6-5 2016 到 2018 年教育支出（对数）变化趋势图

ANOVA 的结果显示，年份和省份的交互作用并不显著 ($F_{2,1255} = 1.724$, $p = 0.179$, $\eta^2 = 0.003$)，这说明从 CFPS 的数据来看，高考改革并不能影响教育支出。

综合 ANOVA 的定量结果和图表的定性呈现，可以发现上海、浙江与其他省份的对数教育支出的均值表现出了一定的交互。相较于其他省份，上海和浙江的对数教育支出的均值都有更高的上涨趋势，但由于上海和浙江的样本变异性较大，结果并不显著。

教育支出是与其他多重因素相关的量，后续本研究也在教育支出中考虑了家庭的财力，如家庭纯收入、储蓄等因素，但结果均不显著，详见附录 B。

6.5.2 情绪指标

选取 CFPS 中问卷填写者报告的学业压力。学业压力取值为 1 至 5，取值越大代表压力越大。删除部分不满足条件的样本后，最终用于分析的样本构成如表 6-5 所示。

<i>year</i>	<i>province</i>	<i>n</i>
2016	Shanghai	14
	Zhejiang	12
	Otherwise	590
2018	Shanghai	19
	Zhejiang	8
	Otherwise	555

表 6-5 情绪指标分析样本构成

同样地，上海和浙江分配到的样本量比较小，需要对结论保持谨慎。

筛选之后，学业压力的分布如图 6-6，分布形态可以开展方差分析。

对学业压力进行二因素方差分析，结果如表 6-6 所示。

从图像上看，2016 年和 2018 年各地区学业压力指标 ($pressure$) 的变化趋势如图 6-7 所示。

ANOVA 的结果显示，年份和省份的交互项并不显著 ($F_{2,1192} = 1.065$, $p = 0.345$, $\eta^2 = 0.002$)，说明从 CFPS 的数据来看，尚且不足以认为高考改革足以对学业压力产生影响。

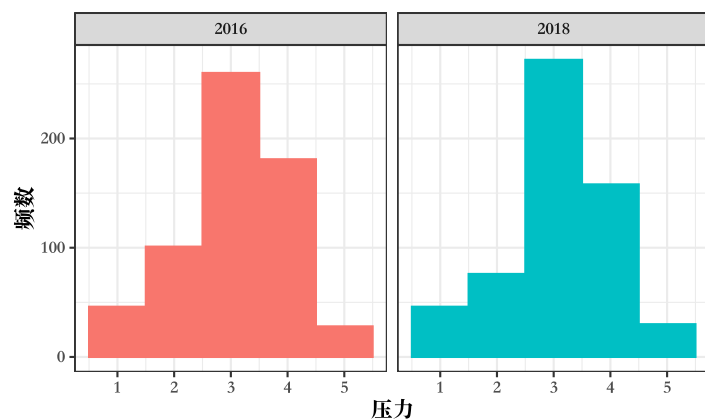


图 6-6 学业压力分布

ANOVA Table							
Dependent Variable: <i>pressure</i>							
Between-Subjects Factor: <i>year, province</i>							
	<i>MS</i>	<i>MSE</i>	<i>df</i> ₁	<i>df</i> ₂	<i>F</i>	<i>p</i>	η^2
<i>year * province</i>	0.984	0.924	2	1192	1.065	0.345	0.002
Levene's Test for Homogeneity of Variance							
	<i>F</i>		<i>df</i> ₁		<i>df</i> ₂		<i>p</i>
DV: <i>pressure</i>	0.772		5		1192		0.570
Descriptive Statistics							
<i>year</i>	<i>province</i>		<i>M</i>		<i>SD</i>		<i>n</i>
2016	Shanghai		2.929		0.917		14
2016	Zhejiang		3.167		1.337		12
2016	Otherwise		3.073		0.961		590
2018	Shanghai		2.526		0.772		19
2018	Zhejiang		2.875		1.126		8
2018	Otherwise		3.108		0.957		555

表 6-6 学业压力的二因素方差分析

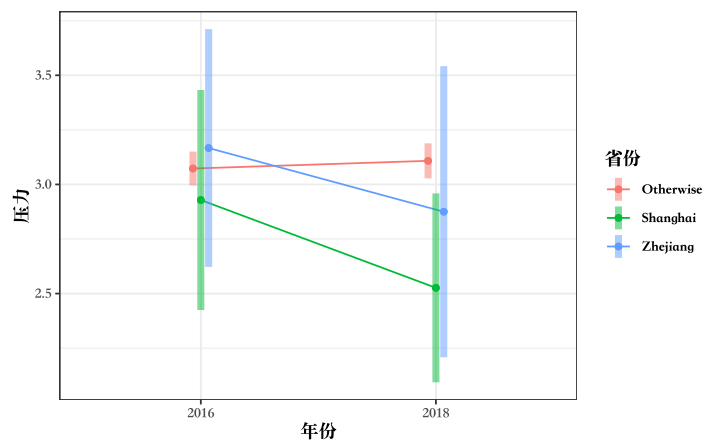


图 6-7 2016 到 2018 年学业压力指标变化趋势

综合 ANOVA 的定量结果和图表的定性呈现,可以发现上海、浙江与其他省份的学业压力的均值表现出了一定的交互。相较于其他省份,上海和浙江的学业压力的均值表现出了下降趋势,但由于上海和浙江的样本变异性较大,结果不显著。

之后,本研究同时选取了 CFPS 中在情绪方面刻画内卷程度的其他指标,但均没有得到显著的结果,详见附录 B。

6.6 对 CFPS 数据的讨论

综合所有得到的结果,可以得出以下结论:

- 1) 从均值变化上看高考改革关于年份对因变量的影响,高考改革地区的学生教育支出有所增加,情绪指标有所改善;
- 2) 从方差分析的结果上看,以上结果均不显著。

上述 ANOVA 分析结果不显著的原因可能来自取样。CFPS 数据库覆盖的上海和浙江的样本较少,缺失值较多,变异性较大,难以代表上海和浙江高中生的整体情况,致使 ANOVA 分析难以获得统计显著的结果。

这启发本研究进一步探索结果不显著的原因究竟是数据集缺陷还是高考改革对考生内卷程度的确无影响。在 CFPS 数据集里,能用于分析高考改革的样本(上海和浙江)过小,因此结论的效力有限;且现阶段便于获取的 CFPS 数据仅至 2018 年,而大批高考改革主要发生在 2020 年及以后,因此,有必要主动向高中生群体发放问卷,获得最新的一手数据,以便进一步的分析。

7 问卷调查

7.1 问卷基本情况

7.1.1 问卷主要内容

在对 CFPS 数据进行分析之后,研究小组设计了直接面向高中生及往届高考生发放的问卷,希望能得到最新的一手数据,切实探索主要发生在 2020 年及以后的高考改革对高中生的影响。

问卷调查的核心目的是评估高考综合改革中引入等级赋分制对考生的投入是否有显著的影响,因此问卷中收集了填写者的高考省份、高考年份、选科情况、学习时间投入、学习时间在主科和选科之间的分配、各科擅长程度、压力、睡眠、作业量、考试频次、家庭支持程度等指标,能更真实全面地获取高中生数据用于分析。其中,学习时间投入、学习时间在主科和选科之间的分配等指标能直接衡量考生的投入;家庭支持程度、压力、睡眠等指标能间接反映考生的内卷程度;各科擅长程度、作业量等属于需要观察并控制的外部变量。

问卷具体设计详见附录 D。

7.1.2 问卷样本分布

结合高考改革的推进情况,该调查向北京、天津、上海、浙江、山东、福建这六个经历过高考改革的地区^①集中发放问卷,同时辐射全国各个省份以提升数据多样性^②。

最终,回收答卷 1196 份,其中有效答卷 757 份^③。数据样本在全国范围的分布如图 7-1 所示。

从样本在全国范围内的分布来看,问卷的发放达到了预期的效果^④。

受访者参加高考的年份的分布如图 7-2 所示。

受访者经历的高考类型的分布如图 7-3 所示。

① 选取这六个省份的考量见第 5 部分。

② 新疆、西藏、青海、宁夏由于没有目前尚未施行高考改革,所以暂不覆盖。香港、澳门、台湾地区的高招模式不同于其他地区,因此也暂不覆盖。

③ 有效答卷的筛选条件详见附录 C。

④ 北京样本 147 个,天津样本 40 个,上海样本 44 个,浙江样本 161 个,山东样本 67 个,福建样本 174 个,其他地区样本共 124 个。

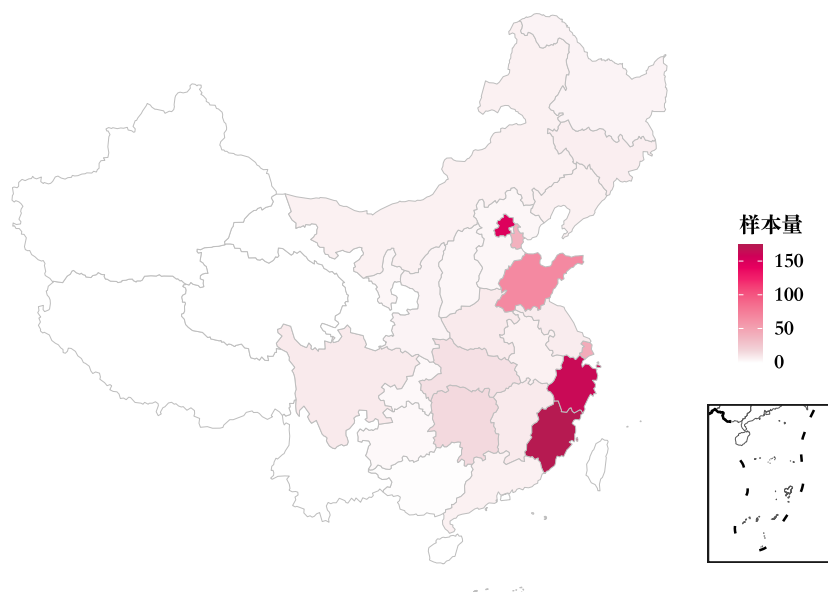


图 7-1 问卷调查样本地域分布

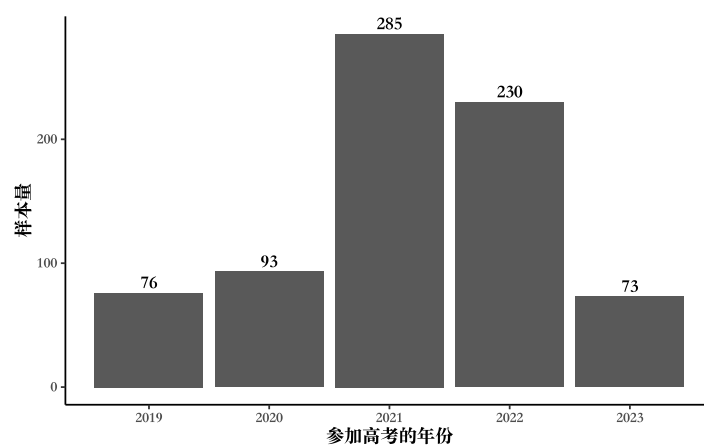


图 7-2 问卷调查样本年份分布

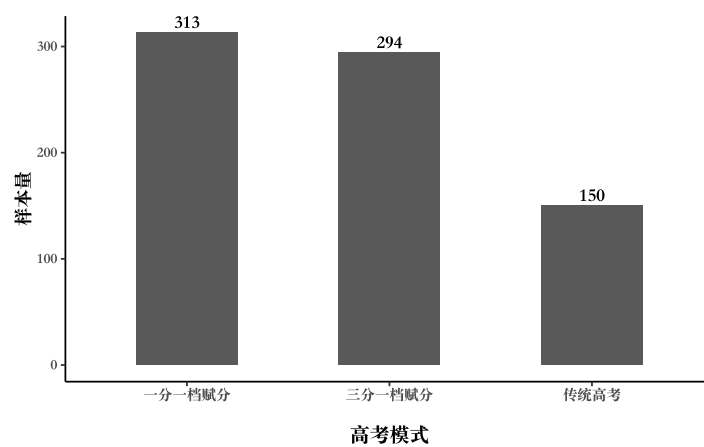


图 7-3 问卷调查样本高考类型分布

7.2 主科和选科时间投入的模型建构及分析

7.2.1 模型变量的确定

结合理论，本研究重点关注高考改革对高中生的学习时间的的影响，因此选择主科投入时间和选科投入时间作为被解释变量，高考改革为解释变量。变量之间的线性相关如图 7-4 所示。在探究变量之间的线性相关关系之后，初步确定建构模型的变量结构。

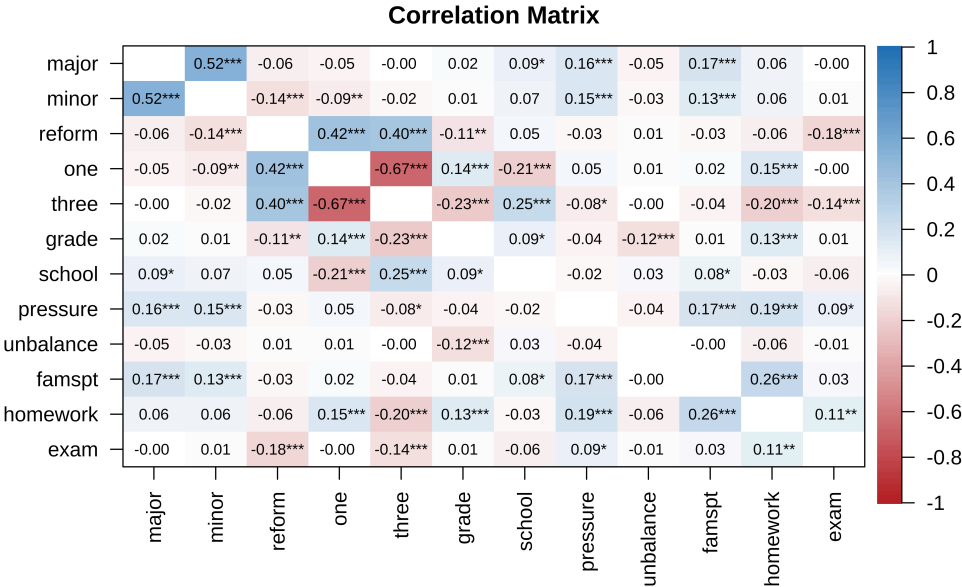


图 7-4 各变量相关程度热力图

被解释变量：主科周末^①日均投入时间（下简称主科投入）和选科周末日均投入时间（下简称选科投入）。

解释变量：高考改革与否或高考改革的种类。具体来说，当区分高考改革与否时，关注考生参加的是否为老高考；而当区分高考改革的种类时，将考生参加的高考的选科计分方式分为原始分、一分一档赋分和三分一档赋分，如图 7-5 所示。

控制变量：高考年份、年级排名、学校实力、偏科程度、学业压力、家庭支持、作业量。社会上的资源随着时代迅猛发展而变得稀缺，年份可能会影响投入；年级排名可能与考生的学习模式与方法有密切关系，因而与投入相关；学校实力决定考生的资源和环境，对投入有显著影响；偏科程度可能与投入模式相关，因此需要控制；作业量可能会挤占考生周末的学习时间，必须加以控制；学业压力、家庭支持程度等作为衡量考生内卷程度的间接指标，可能作用于考生的投入。对于这些变量，本模型予以控制，以明确高考改革对时间投入的偏效应^②。

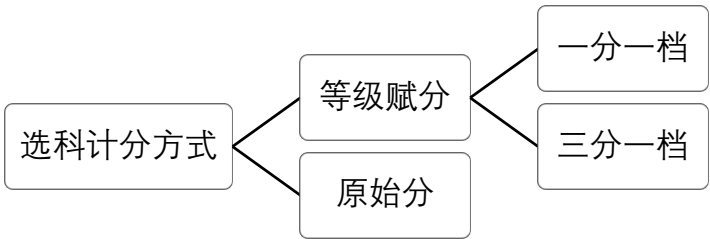


图 7-5 问卷数据分组图示

模型变量的具体定义如表 7-1 所示。

^① 根据现实情况，在高中生上学的日子，他们的学习时间相对固定而无弹性，而休息的日子有更多弹性。因此，休息的日子里的时间利用更能刻画高中生的内卷程度，优先选择问卷填写者汇报的周末日均学习时间进行分析。

^② 经过综合考虑和评估之后，决定不将考试频次纳入回归的自变量。因为受访者对考试频次的回忆不准确，有很多缺失值和极端值。

变量类型	变量名称	符号	定义
被解释变量	主科投入	<i>major</i>	周末日均主科投入（小时）。由受访者的周末学习时间和主科时间分配比 ^① 相乘得出。
	选科投入	<i>minor</i>	周末日均选科投入（小时）。由受访者的周末学习时间和选科时间分配比相乘得出。
解释变量	高考改革	<i>reform</i>	是否经历高考改革。经历高考改革的受访者该变量取值为 1，否则（也即老高考）取值为 0。
	一分一档	<i>one</i>	是否经历一分一档的赋分制度。受访者参加的高考中的选科实行每一阶梯一分的赋分方式时取值为 1，否则为 0。
	三分一档	<i>three</i>	是否经历三分一档的赋分制度。受访者参加的高考中的选科实行每一阶梯三分的赋分方式时取值为 1，否则为 0。
控制变量	高考年份	<i>year</i>	受访者参加高考的年份。
	年级排名	<i>grade</i>	受访者报告的排名在年级的百分位数（上限为 100）。
	学校实力	<i>school</i>	受访者报告的其学校实力在省内的百分位数（上限为 100）。
	偏科程度	<i>unbalance</i>	受访者报告的选科擅长程度（上限为 100）的极差 ^② 。
	学业压力	<i>pressure</i>	受访者报告的其压力在省内同届高中生中的百分位数（上限为 100）。
	家庭支持	<i>famspt</i>	受访者报告的家庭给予的支持程度（上限为 100） ^③ 。
	作业量	<i>homework</i>	受访者报告的所在学校布置的作业量（上限为 100） ^④ 。

表 7-1 问卷调查模型变量的定义

7.2.2 描述性统计

变量名称	符号	最小值	Q_1	中位数	Q_3	最大值	均值	标准差
主科投入	<i>major</i>	0.041667	2.5	4	5.4	16.8	4.152324	2.179607
选科投入	<i>minor</i>	0.041667	2	3.15	4.2	12.25	3.222401	1.854553
年级排名	<i>grade</i>	5	63	82	95	100	76.77675	21.64007
学校实力	<i>school</i>	2	60	80	95	100	75.60621	22.20527
偏科程度	<i>unbalance</i>	0	26	48	61	100	46.0568	24.35736
学业压力	<i>pressure</i>	0	15	25	41	100	30.18626	20.40252
家庭支持	<i>famspt</i>	0	50	71	86	100	66.17702	26.38167
作业量	<i>homework</i>	3	56	73	85	100	69.8996	20.94565

表 7-2 问卷调查主要变量的描述性统计

① 在问卷中，只有收集到主科和选科的时间分配比，并没有区分周末或工作日。此处假定这样的时间分配比在周末和工作日不变。

② 同样地，可以定义主科范围内的偏科和全科范围内的偏科。但经过相关系数的检验发现，三者两两高度线性相关，为防止多重共线性，在模型中只选取选科的偏科程度，并称之为“偏科程度”。

③ 问卷中询问“您认为您的家庭在您高三一年在支持您备考上的投入（如饭菜改善、课外辅导开支、陪伴等）程度为多少”。

④ 问卷中询问“您认为高三一年您的学校布置的作业量如何”。

对于要参与线性回归的连续型变量，描述性统计结果如表 7-2 所示。

各个变量的分布直方图见附录 C。

综合各个变量的描述性统计和分布，可以为研究提供一些初步的观察。比如本研究的研究者来自北京大学，问卷从研究者的社交网络向外扩散，这可能对各个变量的分布产生一定影响，比如年级排名和学校实力中有不少样本集中于上游。这些都提示我们在得出结论时应该更加谨慎。

7.2.3 模型建构及回归方法

确定两类模型。第一类区分高考改革与否，第二类区分高考改革的种类。

第一类：区分高考改革与否

原始模型

$$major = a_0 + a_1 \times reform + \varepsilon \quad (7.1)$$

$$minor = a_0 + a_1 \times reform + \varepsilon \quad (7.2)$$

控制变量模型

$$\begin{aligned} major = & a_0 + a_1 \times reform + a_4 \times year + a_5 \times grade + a_6 \times school \\ & + a_7 \times unbalance + a_8 \times pressure + a_9 \times famspt + a_{10} \times homework + \varepsilon \end{aligned} \quad (7.3)$$

$$\begin{aligned} minor = & a_0 + a_1 \times reform + a_4 \times year + a_5 \times grade + a_6 \times school \\ & + a_7 \times unbalance + a_8 \times pressure + a_9 \times famspt + a_{10} \times homework + \varepsilon \end{aligned} \quad (7.4)$$

第二类：区分高考改革的种类

原始模型

$$major = a_0 + a_2 \times one + a_3 \times three + \varepsilon \quad (7.5)$$

$$minor = a_0 + a_2 \times one + a_3 \times three + \varepsilon \quad (7.6)$$

控制变量模型

$$\begin{aligned} major = & a_0 + a_2 \times one + a_3 \times three + a_4 \times year + a_5 \times grade + a_6 \times school \\ & + a_7 \times unbalance + a_8 \times pressure + a_9 \times famspt + a_{10} \times homework + \varepsilon \end{aligned} \quad (7.7)$$

$$\begin{aligned} minor = & a_0 + a_2 \times one + a_3 \times three + a_4 \times year + a_5 \times grade + a_6 \times school \\ & + a_7 \times unbalance + a_8 \times pressure + a_9 \times famspt + a_{10} \times homework + \varepsilon \end{aligned} \quad (7.8)$$

其中 ε 表示随机扰动，是模型的变量不能解释的部分。

以如上确定的模型为基准模型，剔除异常值^①后进行多元线性回归（OLS）。

7.2.4 结果汇报及分析

主科的基准回归结果如表 7-3 所示。

对主科的回归结果显示，无论是仅区分是否改革，还是将其划分为一分一档和三分一档，高考改革前后考生在主科上的备考时间投入无显著差异。同时，区分高考改革的种类的

^① 剔除异常值的方法是，去除在上确界为 $Q_3 + 1.5IQR$ 和下确界为 $Q_1 - 1.5IQR$ 的区间以外的值，其中 IQR 为四分位距， $IQR = Q_3 - Q_1$ 。经过对各个变量分布的观察，在回归模型中只去除了变量主科投入（*major*）和选科投入（*minor*）的异常值。第七部分中所有剔除异常值的方法均如此。

变量	<i>major</i>			
	(7.1)	(7.3)	(7.5)	(7.7)
<i>reform</i>	-0.1437 (0.1835)	0.1547 (0.2064)		
<i>one</i>			-0.1589 (0.1995)	0.2451 (0.2329)
<i>three</i>			-0.1273 (0.2019)	0.0957 (0.2182)
<i>year</i>		-0.1934** (0.0753)		-0.2132*** (0.0789)
<i>grade</i>		0.0025 (0.0033)		0.0019 (0.0034)
<i>school</i>		0.0088*** (0.0032)		0.0095*** (0.0033)
<i>unbalance</i>		-0.0042 (0.0035)		-0.0043 (0.0035)
<i>pressure</i>		0.0109*** (0.0030)		0.0109*** (0.0030)
<i>famspt</i>		0.0131*** (0.0028)		0.0131*** (0.0028)
<i>homework</i>		0.0009 (0.0036)		0.0004 (0.0036)
截距项	4.1703*** (0.1646)	392.6538** (152.0629)	4.1703*** (0.1647)	432.6842*** (159.4136)
R^2	0.0008	0.0776	0.0009	0.0784
样本量	748	748	748	748

注：*、** 和 *** 分别表示在 0.1、0.05 和 0.01 水平上显著。下同。

表 7-3 基准回归结果（主科）

回归模型中，一分一档与三分一档时考生在主科上的备考时间投入也没有显著差异。这表明高考是否改革、如何改革均不影响考生在主科上的备考时间投入，与理论预测 1 相符。

从回归结果可以看出，年份（*year*）、学校实力（*school*）、学业压力（*pressure*）和家庭支持（*famspt*）均与考生在主科上的备考时间投入显著相关，但这些因素对考生主科上备考的时间投入的解释力仍较弱。

选科的基准回归结果如表 7-4 所示。

对选科的回归结果显示，高考改革后的考生在选科上的备考时间投入显著低于高考改革前的考生，前者周末在选科上的日投入时间约比后者低 36 分钟。区分一分一档与三分一档后，两种改革形式下考生在选科上的备考时间投入均显著低于未改革时，周末的日投入时间分别低 41 分钟和 33 分钟。相较三分一档，参加一分一档改革的高考的考生在选科上的备考时间投入更低，与理论预测 2 一致。

除此之外，学校实力（*school*）、学业压力（*pressure*）和家庭支持（*famspt*）均与考生在选科上的备考时间投入显著相关，但这些因素对考生主科上备考的时间投入的解释力较弱（回归系数较小）。

变量	<i>minor</i>			
	(7.2)	(7.4)	(7.6)	(7.8)
<i>reform</i>	-0.5855*** (0.1529)	-0.5970*** (0.1758)		
<i>one</i>			-0.6709*** (0.1663)	-0.6808*** (0.1981)
<i>three</i>			-0.4946*** (0.1681)	-0.5417*** (0.1859)
<i>year</i>		0.0037 (0.0641)		0.0221 (0.0672)
<i>grade</i>		-0.0027 (0.0029)		-0.0021 (0.0029)
<i>school</i>		0.0084*** (0.0028)		0.0078*** (0.0029)
<i>unbalance</i>		-0.0018 (0.0030)		-0.0017 (0.0030)
<i>pressure</i>		0.0067*** (0.0026)		0.0067*** (0.0026)
<i>famspt</i>		0.0064*** (0.0024)		0.0064*** (0.0024)
<i>homework</i>		0.0021 (0.0030)		0.0026 (0.0031)
截距项	3.5917*** (0.1371)	-5.2062 (129.5748)	3.5917*** (0.1371)	-42.3060 (135.7273)
R^2	0.0194	0.0596	0.0216	0.0607
样本量	745	745	745	745

表 7-4 基准回归结果（选科）

7.3 家庭支持与学业压力的模型建构与分析

7.3.1 模型变量的确定

在 CFPS 数据的实证分析中，家庭教育支出及考生情绪指标均未在高考改革前后体现出显著的变化。本部分将进一步利用问卷数据确定二者变化不显著的原因是 CFPS 数据集本身缺陷，还是高考改革对二者的确无影响。

据此，本部分分别选取家庭支持（*famspt*）和学生学业压力（*pressure*）为被解释变量，是否经历高考改革（*reform*）为解释变量。同时，控制高考年份、年级排名、学校实力、偏科程度、学业压力、作业量等变量。

7.3.2 模型建构

家庭支持

原始模型

$$famspt = a_0 + a_1 \times reform + \varepsilon \quad (7.9)$$

控制变量模型

$$famspt = a_0 + a_1 \times reform + a_2 \times year + a_3 \times grade + a_4 \times school + a_5 \times unbalance + a_6 \times pressure + a_8 \times homework + \varepsilon \quad (7.10)$$

学业压力

原始模型

$$pressure = a_0 + a_1 \times reform + \varepsilon$$

(7.11)

控制变量模型

$$pressure = a_0 + a_1 \times reform + a_2 \times year + a_3 \times grade + a_4 \times school + a_5 \times unbalance + a_7 \times famspt + a_8 \times homework + \varepsilon$$

(7.12)

其中 ε 表示随机扰动，是模型的变量不能解释的部分。
以如上确定的模型为基准模型，剔除异常值后进行多元线性回归（OLS）。

7.3.3 结果汇报及分析

以上模型的回归结果如表 7-5 所示。

变量	<i>famspt</i>		<i>pressure</i>	
	(7.9)	(7.10)	(7.11)	(7.12)
<i>reform</i>	-1.6665 (2.4064)	-2.6008 (2.6865)	-1.7915 (2.2215)	0.1506 (0.2143)
<i>year</i>		1.2021 (0.9861)		-0.0286 (0.0784)
<i>grade</i>		-0.0381 (0.0437)		0.0047 (0.0035)
<i>school</i>		0.1230*** (0.0421)		0.0087** (0.0034)
<i>unbalance</i>		0.0115 (0.0455)		-0.0044 (0.0036)
<i>pressure</i>		0.1262*** (0.0388)		
<i>famspt</i>				-0.0015 (0.0029)
<i>homework</i>		0.2988*** (0.0453)		-0.0018 (0.0037)
截距项	67.5133*** (2.1548)	-2394.8470 (1991.9560)	47.4933*** (1.9892)	63.6860 (158.3844)
R^2	0.0006	0.0919	0.0009	0.0163
样本量	757	757	757	757

表 7-5 家庭支持和学业压力对是否高考改革的回归结果

回归结果显示，无论是否控制变量，家庭支持和学业压力在高考改革前后均无显著变化，与 CFPS 数据的实证分析结果一致。这意味着高考改革对考生的影响仅反映在考生的时间投入上。考生的情绪、心理及非考生本人的因素（如考生家庭支持）等不受其影响。

7.4 睡眠时间的模型建构及分析

7.4.1 模型变量的确定

中华人民共和国教育部于 2021 年 3 月 30 日印发《教育部办公厅关于进一步加强中小学生睡眠管理工作的通知》（以下简称《通知》），强调中小学生睡眠的重要性。《通知》明确，学校和家庭应当合理安排学生的就寝时间，防止学业过重挤占睡眠时间。在高考改革后，考

生在备考上的总投入时间明显下降，但这些省下的时间究竟是如《通知》所希望的那样增加了考生的休息睡眠时间，还是被考生用在了其他休息睡眠以外的地方，仍需进一步研究。

据此，本部分以问卷中收集的考生睡眠时间为因变量，是否经历高考改革为自变量进行回归分析。睡眠时间的变量定义及描述性统计见表 7-6。

变量名称	符号	定义				
睡眠时间	<i>sleep</i>	受访者报告的日均睡眠（含午睡）时间（小时）。				
描述性统计						
最小值	Q_1	中位数	Q_3	最大值	均值	标准差
4	6.5	7	7.542	9.5	7.030	0.8674145

表 7-6 睡眠时间变量的定义及描述性统计

7.4.2 模型建立

基于上述分析，建立如下模型。

原始模型

$$sleep = a_0 + a_1 \times reform + \varepsilon \tag{7.13}$$

控制变量模型

$$sleep = a_0 + a_1 \times reform + a_2 \times year + a_3 \times grade + a_4 \times school + a_5 \times unbalance + a_6 \times pressure + a_7 \times famspt + a_8 \times homework + \varepsilon \tag{7.14}$$

其中 ε 表示随机扰动，是模型的变量不能解释的部分。

以如上确定的模型为基准模型，剔除异常值后进行多元线性回归（OLS）。

7.4.3 结果汇报及分析

以上模型的回归结果如表 7-7 所示。

回归结果表明，高考改革对考生睡眠时间的影响不显著，且其回归系数为负。这与考生从备考中省下时间、理应增加睡眠时间（系数为正）的预期并不一致。可以推断，考生的备考压力减轻后，将省出的时间投入了休息睡眠之外的事情，可能是体育运动、课外阅读、娱乐活动等活动。高考考生睡眠长期处于不足的状态，而高考赋分制改革对内卷程度的减轻并未使考生在睡眠方面受益。

7.5 问卷分析结论

从上述模型对问卷的分析中可以得出以下结论：

- 1) 高考改革降低了考生的内卷程度，且该影响完全反映在考生在选科上投入的备考时间的降低上；高考改革对考生在主科上的时间投入没有显著影响；
- 2) 一分一档制对考生内卷程度的降低显著高于三分一档制；
- 3) 高考改革对考生内卷程度的降低在考生的情绪、心理以及非考生本人因素（如考生家庭支持）上无明显体现；
- 4) 高考改革并未使考生将省下的时间用于睡眠，而是用于休息睡眠以外的活动。

变量	<i>sleep</i>	
	(7.13)	(7.14)
<i>reform</i>	-0.0898 (0.0827)	-0.0524 (0.0933)
<i>year</i>		-0.0436 (0.0342)
<i>grade</i>		0.0028* (0.0015)
<i>school</i>		0.0042*** (0.0015)
<i>unbalance</i>		-0.0019 (0.0016)
<i>pressure</i>		-0.0054*** (0.0014)
<i>famspt</i>		-0.0018 (0.0013)
<i>homework</i>		-0.0038** (0.0016)
截距项	7.1021*** (0.0741)	95.2880 (69.0996)
R^2	0.0017	0.0919
样本量	696	696

表 7-7 睡眠时间对是否经历高考改革的回归结果

8 实验

8.1 实验设计

由于问卷调查的实证分析利用的是历史数据，且本研究所关注的赋分制度与内卷程度的关联可能被错综复杂的其他关系遮盖，譬如尽管模型中控制了偏科程度，但仍然不能保证完全避免高考改革中赋分制以外的因素（如自由选科）的影响，因此需要以控制变量的方法加以验证。同时，之前的数据对象是高中生，而本研究希望赋分制度的有效性和意义在更广泛的对象和情境上得到印证。因此，在对 CFPS 和高中生问卷的数据分析之后，本研究设计并开展了以下实验。

为模拟类似高考的竞争场景，实验中设置了“打字”和“口算”两个“科目”。打字和口算均借助网络平台进行^①。发放给参与者的实验细着重强调参与者可以反复尝试两个项目以反映最高水平。具体而言，单次打字测试为 1 分钟的中文练习，每个参与者面对的任务相同，参与者重复尝试时也呈现同一文段，系统会记录每次打字速度“cpm”（character per minute，每分钟输入汉字数）。实验取单次最快速度作为参与者“打字”科目的最好成绩。单次口算练习是基础时长为 1 分钟的两个两位数相加的计算，每做对一题得一分并加时两秒钟，每做错一题则扣一分，系统会记录单次最高的得分和总尝试次数；实验取单次最高得分作为参与者“口算”科目的最好成绩。

在总计分方式上，本研究始终将“口算”作为原始分科目（即“主科”）直接计入最好成绩，而对“打字”科目实行不同的计分方式以模拟赋分机制，这也符合高考主科和选科的情景。实验招募的参与者将会被随机分到三组，其中一个为控制组，两个为实验组。在控制组中，“打字”科目以原始最好成绩计入（称为原始分组）；在实验组中，对于“打字”的成绩，一组的赋分方式为每档 2 分（称为密赋分组），一组的赋分方式为每档 5 分（称为疏赋

^① 打字借助的平台的链接为 <https://dazi.91xjr.com>；口算借助的平台的链接为 <https://www.preplounge.com/en/mental-math/add/1>。

分组)。疏赋分组和密赋分组统称为赋分组。

密赋分组和疏赋分组的赋分规则如图 8-1 和图 8-2 所示，其中数轴代表排名，轴下字母代表赋分等级，轴上数字代表赋分得分。

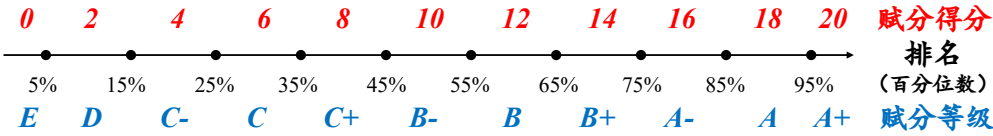


图 8-1 密赋分组赋分规则

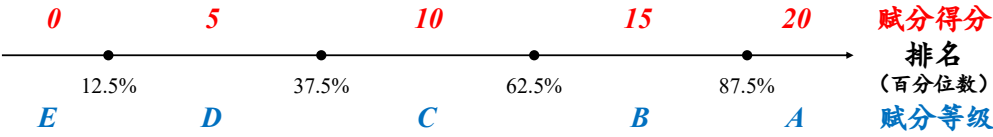


图 8-2 疏赋分组赋分规则

本实验线上开展。在收到实验细则后^①，参与者拥有一周左右的时间自行实验，最终通过指定的问卷提交打字和口算的分别的尝试次数和最佳成绩^②。

所有的参与者都会收到 3 元的基础实验费。在此基础上，根据同一分组内两个科目总分的排名情况，部分参与者还将获得额外的奖励。具体而言，排名在前 2% 的参与者会获得额外 100 元的奖励，排名在 2% 至 10% 的参与者能得到额外 20 元的奖励，10% 及以后的参与者没有额外奖励（参与者未被告知与实验人数相关的信息）。设置奖励在不同梯度之间悬殊的差异，目的是仿照高考录取中不同等级院校对于考生的巨大落差，从而形成贴近现实的激励。

8.2 参与者情况

本实验一共有报名者 255 人，最终实际参与者 117 人，其中原始分组 42 人，密赋分组 39 人，疏赋分组 36 人。参与者绝大部分都是大学本科在读学生，也有少量高中在读学生和大学毕业生。参与者的地域分布如图 8-3 所示。

在参与者招募的过程中，由于招募信息和报名问卷主要依赖于人际关系传播，导致参与者同质性高，且出现地域性的集中趋势，因此对于部分结论应保持谨慎态度。

8.3 实验变量

结合实验的目的与分组，本实验数据的分组变量有两个划分层级。首先，可以按照是否进行赋分来划分，分为原始分组和赋分组；接着，可以按照赋分的类型进一步划分，分为原始分组、密赋分组和疏赋分组，如图 8-4 所示。

对于因变量，选择总投入时间作为衡量内卷程度的主要指标。由于口算和打字的尝试次数不具有直接的可比性和可加性，选用总投入时间刻画内卷更为合理。本部分重点关注总投入时间，并且细分到口算和打字两个科目进行分析。

综上所述，实验中的变量具体定义如表 8-1 所示。

^① 通过邮件形式向参与者发放实验指导手册，手册内容包括赋分制度、奖励设置、一道理解性测试题、回收结果用的问卷链接、一份详尽打字和口算网站的使用指南，并附上实验者的联系方式以便答疑。在理解性测试题中，给出一组由最快打字速度、最快打字速度排名、最高口算得分组成的数据，让参与者按照手册中讲解的赋分方式计算综合得分并在问卷中填写答案，以便于数据筛选。不同组别的实验指导手册之间，除了赋分制度和相应的理解性测试题数据有所不同之外，其余内容完全一致。详见附录 E。

^② 实验同时要求参与者提交网页截图，网页截图显示尝试次数和最佳成绩的信息，以确保报告的真实性。打字网站记录了参与者每一次的尝试和成绩，口算网站记录了参与者尝试的总次数和最好成绩，两者均不可取消和更改。

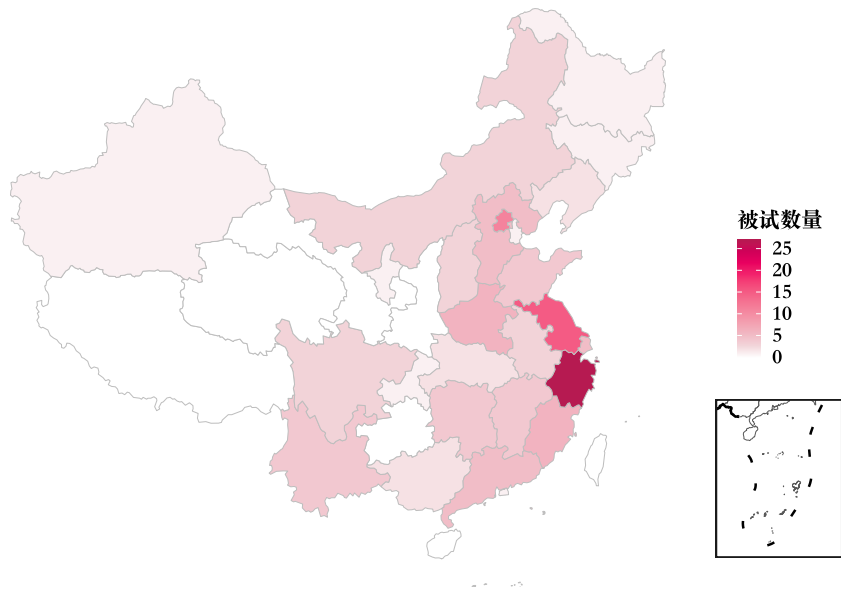


图 8-3 全国范围内的实验参与者分布

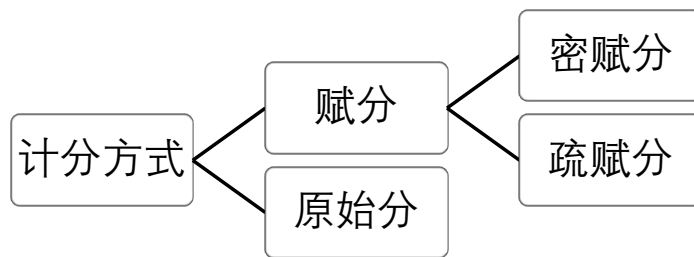


图 8-4 实验分组图示

变量类型	变量名称	符号	定义
因变量	总投入时间	$time_{total}$	参与者在口算和打字上花费的总时间。
	口算投入时间	$time_{calculate}$	参与者在口算上花费的时间（估计值） ^① 。
	打字投入时间	$time_{type}$	参与者在打字上花费的时间 ^② 。
	口算尝试次数	$try_{calculate}$	参与者在口算上尝试次数。
	打字尝试次数	try_{type}	参与者在打字上尝试次数。
分组变量	计分类型	$group$	该变量对于原始分组的参与者取值为 Control，对于密赋分组的参与者取值为 Dense，对于疏赋分组的参与者取值为 Sparse。

表 8-1 实验变量的具体定义

8.4 实验结果

在实验中，参与者可以在打字和口算网站上进行无限次打字或口算的尝试；而由于每个人的打字能力、口算水平相对固定，在参与者熟悉网页操作方式后，成绩并不会随着尝试

^① 口算网站不能记录参与者的口算时间，由于口算的每次尝试用时不一致，此处只能采用估计值。具体来说，在口算的一次尝试中，答对一道题即可增加两秒钟的作答时间，因此可以由口算最佳成绩倒推参与者最佳尝试时所花费的时间，并进一步结合尝试次数给出在口算上的尝试时间的估计值。

^② 同样地，打字网站不能记录参与者的打字时间，但由于打字的每次尝试均用时一分钟，因此可以得出准确的时间。

次数的增多而有显著提升。如图 8-5 和图 8-6 所示，以尝试次数的中位数为界，在右侧的散点没有呈现出尝试次数和最佳成绩间明显的相关性，进一步经检验得出相关系数不显著^①。

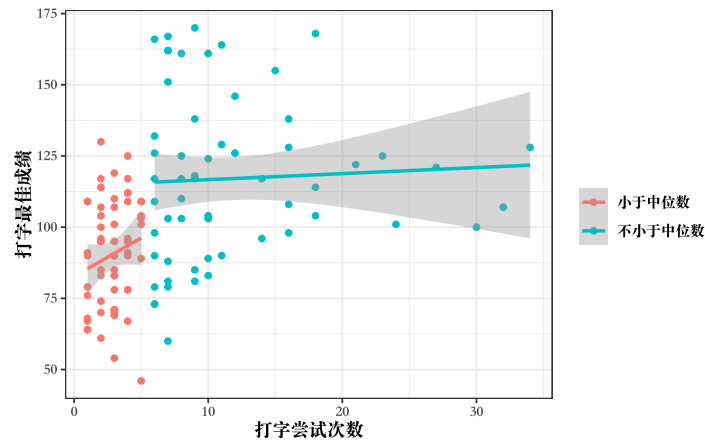


图 8-5 打字尝试次数和最佳成绩的散点图^②

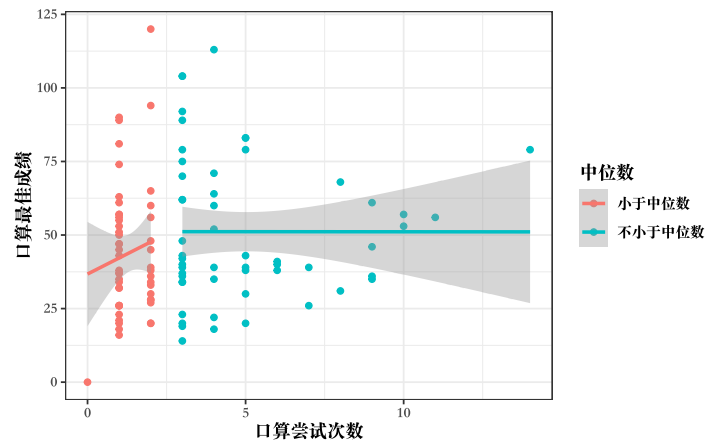


图 8-6 口算尝试次数和最佳成绩的散点图^③

打字和口算的成绩对参与者而言，最初的尝试由于熟悉环境和规则等因素有助于成绩提高，而后续的甚至过量的尝试并无益于提高成绩。过量尝试次数与成绩无关，可见本实验成功激发了参与者的内卷行为。

变量	组别 ^④ 1	组别 2	<i>t</i>	<i>p</i>	显著性
<i>time_{type}</i>	Control	Dense	2.007502	0.048	**
	Control	Sparse	1.076461	0.285	
	Dense	Sparse	-0.838470	0.405	

表 8-2 打字投入时间的分组 *t* 检验

① 本部分希望通过打字和口算的最佳成绩反映参与者练习后达到的最终实力，而离群值可能包含较大的运气因素。为排除离群值对相关系数显著性的影响，此处剔除了处于 $Q_3 + 1.5IQR$ 和 $Q_1 - 1.5IQR$ 的值，其中 $IQR = Q_3 - Q_1$ 。
 ② 在中位数之后有 $r_{pearson} = 0.051$, $p = 0.702$ 和 $r_{spearman} = 0.160$, $p = 0.231$ 。
 ③ 在中位数之后有 $r_{pearson} = -0.001$, $p = 0.996$ 和 $r_{spearman} = -0.013$, $p = 0.925$ 。
 ④ 组别指 *group* 的取值。后同。

在实验中，打字是赋分科目，对应前文提到的“选科”，也同时是本研究重点关注的会受计分方式影响的变量。对于不同的计分方式， t 检验的结果如表 8-2 所示^①。各组的描述性统计如表 8-3 所示。

变量	group	均值	标准差	样本量
$time_{type}$	Control	9	7.55832	40
	Dense	6.11111	4.39336	36
	Sparse	7.21212	6.40682	33

表 8-3 打字投入时间的分组描述性统计

t 检验的结果表明，对于打字科目，采用原始分和赋分的计分方式能在投入时间的均值上体现出比较明显的差异，赋分组的投入小于原始分组。进一步 t 检验的结果验证了这一点，结果显著 ($t_{107} = 1.888836, p = 0.062$)。并且，差异尤其体现在了原始分组和密赋分组上， t 检验的结果显著 ($t_{74} = 2.007502, p = 0.048$)。打字投入时间的箱型图如图 8-7 所示。

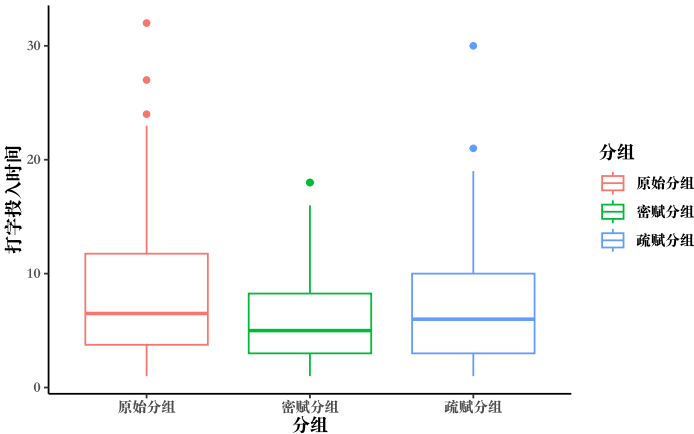


图 8-7 打字投入时间箱型图

在实验中的口算是原始分科目，对应前文提到的“主科”，而主科投入在本研究中被预测并证实了与计分方式无关。对于不同的计分方式， t 检验的结果如表 8-4 所示。

变量	组别 1	组别 2	t	p	显著性
$time_{calculate}$	Control	Dense	1.453035	0.150	
	Control	Sparse	-0.545755	0.587	
	Dense	Sparse	-1.801108	0.076	*

表 8-4 口算投入时间的分组 t 检验

各组的描述性统计如表 8-5 所示。

变量	group	均值	标准差	样本量
$time_{type}$	Control	6.26500	5.50167	40
	Dense	4.80648	2.57596	36
	Sparse	7.06566	7.03273	33

表 8-5 口算投入时间的分组描述性统计

^① 经过 Levene’s Test，得出各组数据的方差同质，因此采用同方差的独立样本 t 检验。实验部分后续的 t 检验也同理，各组数据都通过了方差同质性检验，因此也都采用同方差的独立样本 t 检验。进行 t 检验时去除了总投入时间排名前后各 3% 的极端样本。后同。

t 检验的结果表明，对于口算这一“主科”，原始分、密赋分、疏赋分三个组在口算投入时间上没有体现出明显的差异。疏赋分组和密赋分组差异边缘显著可能是由于参与者并不完全理性。同时，理论模型中，主科和选科达到目标原始分所需的时间投入相互独立的假设在主科和选科任务都较为轻松时可能不再成立，这也可能造成以上 t 检验中的边缘显著。如果进一步探查原始分组和赋分组的参与者之间的投入时间差异， t 检验并不显著 ($t_{107} = 0.354379$, $p = 0.724$)。这些结果印证了“主科投入不随赋分制度改变”的假设。口算投入时间的箱型图如图 8-8 所示。

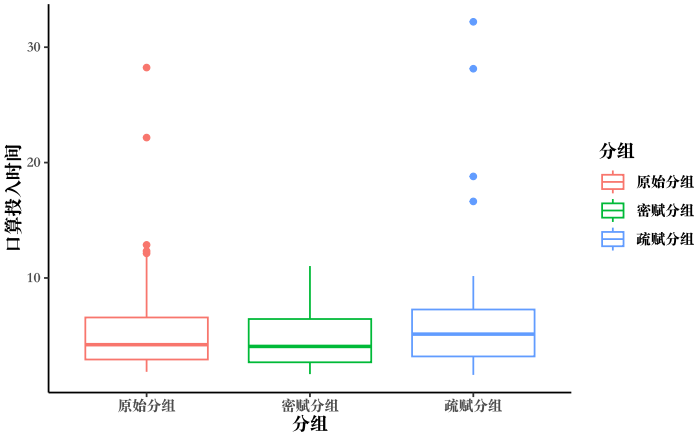


图 8-8 口算投入时间箱型图

最后，在打字和口算两个科目的总投入时间也是对内卷的一种刻画。各组的 t 检验结果如表 8-6 所示。

变量	组别 1	组别 2	t	p	显著性
$time_{total}$	Control	Dense	2.097766	0.039	**
	Control	Sparse	0.389022	0.698	
	Dense	Sparse	-1.642702	0.105	

表 8-6 总投入时间的分组 t 检验

各组的描述性统计如表 8-7 所示。

变量	group	均值	标准差	样本量
$time_{type}$	Control	15.26500	11.01452	40
	Dense	10.91759	6.07183	36
	Sparse	14.27778	10.51243	33

表 8-7 总投入时间的分组描述性统计

t 检验的结果表明，原始分组和赋分组的总投入时间均值有显著差异，赋分组的小于原始分组。并且，差异尤其体现在了原始分组和密赋分组上， t 检验的结果显著 ($t_{74} = 2.097766$, $p = 0.039$)。总投入时间的箱型图如图 8-9 所示。

本实验的规则和过程较为复杂，参与者较少。这为统计上获得显著的结论提高了门槛。尽管有现实条件的约束，以上三个因变量的均值差异基本符合理论预测和此前的实证证明，并且 p 值上也基本呈现了一致的结果。值得指出的是，总投入时间等于“主科”口算投入时间和“选科”打字投入时间的加总，而“主科”口算投入时间已被证实不随计分方式变化，“选科”打字投入时间随计分方式出现了预测之内的变化。理论上，总投入时间的变化应该与“选科”打字投入时间相近，而实验数据也再次印证了这一点，这不仅加强了本研究的结论，而且增强了结论的稳健性。

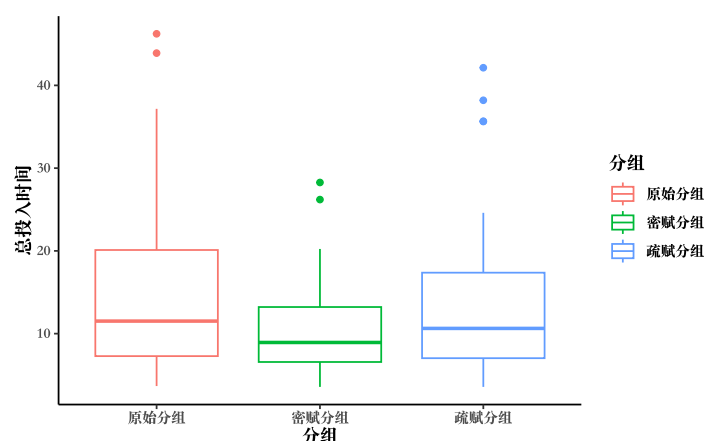


图 8-9 总投入时间箱型图

8.5 实验总结与讨论

本研究的实验在均值趋势上印证了理论部分的预测，也与 CFPS、高中生问卷的实证分析结果相吻合——原始分组的投入程度高于赋分组，密赋分组的投入程度高于疏赋分组。同时这些结论也基本得到了假设检验结果的支持。理论、实证和实验三部分之间的一致性，以及实验中众多指标间的一致趋势，都让本研究的结论更严谨而有力。由于显著性门槛较高，实验数据分析中显著的结果为研究结论提供了有力的支持，同时也为未来的进一步研究探索打开了大门。

但是，实验数据分析中部分趋势性结果的显著性仍不够理想。在数据上，这表现为样本量较小、变异性较大；在实验设计上，这表现为组别之间的差异可以再放大；在操作细节上，这可能与依赖关系网络招募参与者的方式、实验平台与规则的设置都有关。实验中的不足之处值得反思，也为后续研究提供了参考。

9 结论与展望

本研究从对高考等级赋分制的朴素思考出发，参考锦标赛理论、大型竞赛概念，通过模型建构得出理论预测，并在 CFPS 数据库分析、问卷调研和实验中取得了符合理论分析的结果。这些实证分析构成了支持理论模型的证据。

本研究最核心的结论为等级赋分制会减少高考考生在选科上的备考时间投入，同时不增加其主科上的备考时间投入，因此降低总体时间投入，从而在高校录取名额数一定的情况下降低整体内卷程度，保障广大考生的福利。等级赋分制的密度越大，上述影响越明显。此外，本研究也发现引入等级赋分制对考生内卷程度的影响几乎不在考生的情绪状态上反映，也不在非考生本人的考生家庭的支持投入上反映。同时，等级赋分制的引入对考生备考时间投入的降低并未增加考生的睡眠时间。虽然高考考生普遍缺乏睡眠，但他们仍会将省下的时间大量投入非应试学习的其他活动中，这说明对于考生而言，休息固然重要，但对生活丰富、精神满足的需求同样不可忽视。在政策制定中，应充分考虑这一点。

尽管研究在样本结构上不够完美，但倘若研究成果能够经过后续研究者的验证与拓展，或许可以有在以下多种场景广泛应用的可能：在小学、中学、大学的教育评级体系广泛引入等级赋分制，在保证排名不变的同时降低学生内卷程度，给予学生更多自主性，促进学生多元化发展、健康成长；在选拔性考试中体现教育重心，重点科目采用原始分而非等级赋分制计分；乃至将研究成果应用到工作场景的激励机制设计中，用与排名无关的绝对绩效指标或者较为稀疏的等级评定作为激励机制的标准，增强员工的工作动力，从而保障激励效果与员工业绩。

当然，本研究尚存在一定局限，在未来研究中应当予以关注。首先，实证研究上，实验的参与者都是研究小组通过人际网络传播招募的，存在人情因素，本研究设计的激励机制难以充分作用在参与者行为上。参与者覆盖范围也较为局限，倘若收集的参与者样本能够更加具有一般性，问卷调查和实验则可能获得更有代表性的结果，指向更加清晰和稳健的结论。其次，对于问卷调查而言，2021 年 7 月，中共中央启动“双减”政策，虽然只针对

义务教育阶段，对本研究关于高考考生的结论影响有限；但由于双减对教培行业整体产生了不小的冲击，高中生课外补习也受到影响——因此，双减政策本身具有降低内卷的性质，研究不能完全剔除其在实证分析中的影响，后续研究或可对此进行调整。此外，赋分制对内卷的影响是否会和考生个人禀赋相关，是否和考生接受到的信息相关^①等等，都属于值得深度探究的问题。最后，随着高考改革的持续推进，可以期待更多的省份转为赋分制，更多的考生参与到调查中，提供更加丰富的样本，检验研究结论的稳健性。

10 致谢

写到这里，为期五个多月的研究也已接近尾声。

作为一群大二学生首次尝试完整经济学范式学术研究的成果，由于我们知之不多、经验甚少，难免会有很多考虑不周的地方，在此首先衷心感谢导师翁翕老师的细心指导，从一开始选题的确定，到关键变量的定义，再到实证方案的推进和理论模型的建议；无论是线下办公室里的详谈，还是线上视频会议的讨论，在研究的每一个环节，都有翁老师倾囊相授的身影。特别是当开展实证第一阶段的研究时，我们大费周章地处理零散杂乱的数据，得到的结果却十分不理想。在调试数日无果后，我们对二手数据丧失信心，苦恼地向翁老师汇报当前进度，并询问是否需要直接放弃现实数据而全面转向实验。是翁老师鼓励了我们，言明即使是不显著的数据也有其存在的意义，顺着这个思路换种方式继续下去可能会有意外之喜。这无疑极大地鼓舞了我们，并激发了我们的斗志和热情，随后我们仔细分析了二手数据存在的问题，改用自行发放针对性问卷的方式推进，果然取得了不错的效果。可以说，如今我们能拥有如此饱满的实证成果，离不开翁老师当时给予我们的信心。我们也特别感谢所有提供建议和指导的老师和学长学姐们，如为理论模型和文献参考提供思路的刘烁老师，问卷总体思路、听取我们中期检查展示并给予支持和改进建议的涂艳苹和吴祁老师，在实验设计时帮助我们明确实验规范的黄以乐学长，还有提出值得关注并排除双减政策影响的吴楚凡学长。各位学识渊博的老师和学长学姐也让我们见识到了对学术应有的热爱与追求，以及笃行在学术道路时严肃认真、孜孜不倦、精益求精的态度和精神。

其次，感谢所有参与我们实证研究（问卷调查和线上实验）的来自世界各地的朋友们，特别是实验参与者，他们需要投入较长时间和精力，在本就内卷的大环境下“无意识地被迫”投入到另一份“内卷”中——但好在，这一次他们有了更多的选择权，而不用承担太多风险。由于我们经费甚少、自负盈亏，问卷和实验的报酬和奖金相对不多，但他们仍热心参与并展示自身真实的想法和行动，每一份回收的结果都为我们最后的数据分析作出了重要贡献。以及感谢所有帮我们扩散问卷、实验招募通知的好友，我们也为寒假期间在各种社交软件、匿名交流平台上的集中刷屏致歉。

最后，感谢我们团队中的每一位成员，在近半年的时间里，我们有序规划，稳步推进，分工协作，各自发挥所长，展示出出色的组织能力、联络能力、数理能力和统计能力，合力形成了本研究。我们几乎保持了每周一次的固定会议（即使在新冠肆虐、甲流爆发，小组成员不同程度感染时也不例外），在讨论与本研究相关的话题之外，也会分享自己的生活趣事。可以说，通过本次合作，我们既坚定了学术志趣，也收获了珍贵的友谊。

某一个普通个体置身内卷的洪流中可能显得微不足道，其只能随浪潮起起伏伏而无法掌控自我的命运；但如果一个机制能普遍改变一大群人的行为，大家或许就会少一点力不从心的无可奈何之感。作为在“千军万马过独木桥”的高考中侥幸获得优势的幸存者，我们这些北大学子更应去思考这样一个问题——究竟是什么造成了我们常常把“卷死了”挂在嘴边，明知没什么用却还做着“卷”的无用功？导致这个现象的原因是什么？这其中我们可以改变什么？怀着力图改善学弟学妹福祉、或者至少能对我们亲历过的“内卷”进行反思的初衷，我们从无到有、一步步推进研究；即使最终的选题关键词有所更改、研究中包含的内容和原先的预设有所不同，但一路走来，我们始终初心不改，努力向前。感谢在探索高考制度与教育内卷这程旅途中给我们提供思考和指引的所有前辈们，我们或许与他们素未谋面，但他们在某一领域的奠基之作着实给了我们巨大的启发；我们期待在未来能进一步拓展深化相关话题，也欢迎所有对此感兴趣的学者们一起深耕经济学视角下教育内卷这片沃土。

^① 这里指考生对于赋分政策的了解以及理解程度，以及考生所处环境的模拟考试次数以及模拟考试赋分样本池在多大程度上能传达考生在全省的排名信息。

参考文献

- 陈逸清. 教育公平视角下我国高考招生制度改革研究 [D]. 上海师范大学,2021.
- 程伟, 王雨欣. 新高考选考科目赋分方式的比较分析 [J]. 河北师范大学学报 (教育科学版), 2020,22(01):28-32.
- 代逸生, 林科. 基于锦标赛理论视角的高管薪酬差距与公司绩效关系研究——基于房地产行业的实证研究 [J]. 江苏科技大学学报 (自然科学版),2013,27(02):177-182.
- 国务院关于深化考试招生制度改革的实施意见 [J]. 浙江省人民政府公报,2014(34):24-29.
- 黄海龙. 改革赋分方案实现高考评价科学与录取公平 [J]. 教育教学论坛,2020(31):7-10.
- 李付鹏, 宋吉祥, 杜海燕. 新高考学业水平考试两种等级赋分方式的比较研究 [J]. 教育测量与评价,2019(11):14-19.
- 李金波. 新高考招生制度下的等级赋分制 [J]. 教育测量与评价 (理论版),2016(04):48-51.
- 李金波. 高考选考科目等级赋分制的再思考 [J]. 教育评论,2019(03):19-22.
- 罗立祝. 新高考选考科目赋分方案再思考 [J]. 河北师范大学学报 (教育科学版),2020,22(04):14-20.
- 马秋敏. 政治锦标赛下基层干部晋升机制存在的问题及对策研究 [D]. 湘潭大学,2019.
- 魏光兴, 陈永恒. 考虑异质性的锦标竞赛机制设计 [J]. 江西师范大学学报 (自然科学版),2022,46(01):18-24.
- 温忠麟. 新高考选考科目计分方式探讨 [J]. 中国考试, 2017(12):23-29.
- 杨华. 县域晋升锦标赛模式有效性分析 [J]. 长白学刊,2019(05):86-91.
- 于涵, 韩宁, 关丹丹, 章建石, 焦丽亚. 关于改进新高考选考科目赋分方案的若干思考 [J]. 中国高教研究,2018(06):44-49.
- 余静. 薪酬激励对创新产出影响的实验研究 [D]. 南京财经大学,2020.
- 臧铁军, 杨君. 新高考中学业水平考试成绩转换研究 [J]. 教育研究,2017,38(12):68-75.
- 郑若玲, 凌磊, 吴根洲等. “新时代高考综合改革的纵深推进” 笔谈 [J]. 福建师范大学学报 (哲学社会科学版),2020,No.223(04):117-133.
- 周黎安, 李宏彬, 陈烨. 相对绩效考核: 关于中国地方晋升的一项经验研究 [J]. 经济学报第 1 期, 2005.
- 周黎安. 中国地方官员的晋升锦标赛模式研究 [J]. 经济研究,2007(07):36-50.
- Budde J. Bonus pools, limited liability, and tournaments[R]. SFB/TR 15 Discussion Paper, 2007.O'keeffe M, Viscusi W K, Zeckhauser R J. Economic contests: Comparative reward schemes[J]. Journal of labor economics, 1984, 2(1): 27-56.
- Bodoh-Creed A L, Hickman B R. College assignment as a large contest[J]. Journal of Economic Theory, 2018, 175: 88-126.
- Dickinson D L, Isaac R M. Absolute and relative rewards for individuals in team production[J]. Managerial and decision economics, 1998, 19(4-5): 299-310.
- Fehr E, Falk A. Psychological foundations of incentives[J]. European economic review, 2002, 46(4-5): 687-724.
- Geertz C. Agricultural involution: The processes of ecological change in Indonesia[M]. Univ of California Press, 1963.
- Gneezy U, Rustichini A. Pay enough or don't pay at all[J]. The Quarterly journal of economics, 2000, 115(3): 791-810.
- Green J R, Stokey N L. A comparison of tournaments and contracts[J]. Journal of Political Economy, 1983, 91(3): 349-364.
- Harbring C, Irlenbusch B. An experimental study on tournament design[J]. Labour Economics, 2003, 10(4): 443-464.
- Heneman R L, Von Hippel C. Balancing group and individual rewards: Rewarding individual contributions to the team[J]. Compensation & Benefits Review, 1995, 27(4): 63-68.
- Irlenbusch B, Ruchala G K. Relative rewards within team-based compensation[J]. Labour economics, 2008, 15(2): 141-167.
- Lazear E P, Rosen S. Rank-order tournaments as optimum labor contracts[J]. Journal of political Economy, 1981, 89(5): 841-864.
- Kräkel M. Helping and sabotaging in tournaments[J]. International Game Theory Review, 2005, 7(02): 211-228.

- Krishna K, Lychagin S, Olszewski W, et al. Pareto Improvements in the Contest for College Admissions[R]. National Bureau of Economic Research, 2022.
- Olszewski W, Siegel R. Large contests[J]. *Econometrica*, 2016, 84(2): 835-854.
- Olszewski W, Siegel R. Bid caps in large contests[J]. *Games and Economic Behavior*, 2019, 115: 101-112.
- Olszewski W, Siegel R. Performance-maximizing large contests[J]. *Theoretical Economics*, 2020, 15(1): 57-88.
- Orrison A, Schotter A, Weigelt K. Multiperson tournaments: An experimental examination[J]. *Management Science*, 2004, 50(2): 268-279.
- Rosen S. Prizes and incentives in elimination tournaments[J]. 1985.
- Tsoulouhas T, Marinakis K. Tournaments with ex post heterogeneous agents[J]. Available at SSRN 1026073, 2007.

附录 A 命题证明

命题 2 证明:

只需证明 $\int_0^1 g_1(x)y_n(x)f(x)dx$ (下称“前者”) 关于 n 单调递减, $\int_0^1 \frac{f(x)}{g_2(x)}dx \int_0^x y_n(\tilde{x})d\tilde{x}$ (下称“后者”) 关于 n 单调递增。

对于前者, 相邻两项做差有:

$$\begin{aligned}\Delta_{\text{前}} &= \int_0^1 g_1(x)y_n(x)f(x)dx - \int_0^1 g_1(x)y_{n+1}(x)f(x)dx \\ &= \int_0^1 g_1(G_n^{-1}(F(x)) - G_{n+1}^{-1}(F(x)))f(x)dx \\ &= \int_0^1 g_1(F^{-1}(x))(G_n^{-1}(x) - G_{n+1}^{-1}(x))dx\end{aligned}\tag{A.1}$$

记 $h(x) = g_1(F^{-1}(x))$, 则 $h(x) > 0$ 且关于 x 单调递增。

$$\begin{aligned}\Delta_{\text{前}} &= \int_0^1 h(x)(G_n^{-1}(x) - G_{n+1}^{-1}(x))dx \\ &= \sum_{i=1}^{n-1} \int_{\frac{i}{n+1}}^{\frac{i}{n}} h(x)\beta\left(\frac{i-1}{n-1} - \frac{i}{n}\right)dx + \sum_{i=1}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n+1}} h(x)\beta\left(\frac{i}{n-1} - \frac{i}{n}\right)dx \\ &\geq \sum_{i=1}^{n-1} \int_{\frac{i}{n+1}}^{\frac{i}{n}} h\left(\frac{i}{n}\right)\beta\left(\frac{i-1}{n-1} - \frac{i}{n}\right)dx + \sum_{i=1}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n+1}} h\left(\frac{i}{n}\right)\beta\left(\frac{i}{n-1} - \frac{i}{n}\right)dx \\ &= \beta \sum_{i=1}^{n-1} h\left(\frac{i}{n}\right) \left(\left(\frac{i-1}{n-1} - \frac{i}{n}\right) \left(\frac{i}{n} - \frac{i}{n+1}\right) + \left(\frac{i}{n-1} - \frac{i}{n}\right) \left(\frac{i+1}{n+1} - \frac{i}{n}\right) \right) \\ &= \beta \sum_{i=1}^{n-1} h\left(\frac{i}{n}\right) \left(\frac{i(i-n)}{(n-1)n^2(n+1)} + \frac{i(n-i)}{(n-1)n^2(n+1)} \right) \\ &= 0\end{aligned}\tag{A.2}$$

故前者关于 n 单调递减。

对于后者, 相邻两项做差有:

$$\begin{aligned}\Delta_{\text{后}} &= \int_0^1 \frac{f(x)}{g_2(x)}dx \int_0^x y_n(\tilde{x})d\tilde{x} - \int_0^1 \frac{f(x)}{g_2(x)}dx \int_0^x y_{n+1}(\tilde{x})d\tilde{x} \\ &= \int_0^1 \frac{f(x)}{g_2(x)}dx \int_0^x (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))d\tilde{x}\end{aligned}\tag{A.3}$$

只需证明 $\int_0^x (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))d\tilde{x} \leq 0$ 对任意 $x \in [0, 1]$ 成立。

对于任意的 $n \geq 2$, G_n^{-1} 关于 $(\frac{1}{2}, \frac{\beta}{2})$ 对称, F 关于 $(\frac{1}{2}, \frac{1}{2})$ 对称, 故 y_n 关于 $(\frac{1}{2}, \frac{\beta}{2})$ 对称。因而有 $\forall a \in [0, \frac{1}{2}]$:

$$\begin{aligned}
& \int_{\frac{1}{2}-a}^{\frac{1}{2}+a} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} + \int_{\frac{1}{2}}^{\frac{1}{2}+a} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(1-\tilde{x}) - y_{n+1}(1-\tilde{x})) d\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} ((\beta - y_n(\tilde{x})) - (\beta - y_{n+1}(\tilde{x}))) d\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_{n+1}(\tilde{x}) - y_n(\tilde{x})) d\tilde{x} \\
&= 0
\end{aligned} \tag{A.4}$$

故只需考虑 $x \in [0, \frac{1}{2}]$ 的情况。

$\forall i \in \{1, 2, \dots, n-1\}$, 注意到当 $\tilde{x} \in \left(F^{-1}\left(\frac{i}{n+1}\right), F^{-1}\left(\frac{i}{n}\right)\right)$ 时, 有:

$$y_n(\tilde{x}) - y_{n+1}(\tilde{x}) = \beta \cdot \left(\frac{i-1}{n-1} - \frac{i}{n}\right) < 0 \tag{A.5}$$

而当 $\tilde{x} \in \left(F^{-1}\left(\frac{i}{n}\right), F^{-1}\left(\frac{i+1}{n+1}\right)\right)$ 时, 有:

$$y_n(\tilde{x}) - y_{n+1}(\tilde{x}) = \beta \cdot \left(\frac{i}{n-1} - \frac{i}{n}\right) > 0 \tag{A.6}$$

故只需证明 $\forall i \in \{1, 2, \dots, n-1\}$ 且 $F^{-1}\left(\frac{i+1}{n+1}\right) \leq \frac{1}{2}$, 有:

$$\int_{F^{-1}\left(\frac{i}{n+1}\right)}^{F^{-1}\left(\frac{i}{n}\right)} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} + \int_{F^{-1}\left(\frac{i}{n}\right)}^{F^{-1}\left(\frac{i+1}{n+1}\right)} (y_n(\tilde{x}) - y_{n+1}(\tilde{x})) d\tilde{x} \leq 0 \tag{A.7}$$

对 (A.7) 左端变形有:

$$\begin{aligned}
LHS &= \left(F^{-1}\left(\frac{i}{n}\right) - F^{-1}\left(\frac{i}{n+1}\right)\right) \beta \left(\frac{i-1}{n-1} - \frac{i}{n}\right) \\
&\quad + \left(F^{-1}\left(\frac{i+1}{n+1}\right) - F^{-1}\left(\frac{i}{n}\right)\right) \beta \left(\frac{i}{n-1} - \frac{i}{n}\right) \\
&= \frac{\beta}{n(n-1)} \left(iF^{-1}\left(\frac{i+1}{n+1}\right) + (n-i)F^{-1}\left(\frac{i}{n+1}\right) - nF^{-1}\left(\frac{i}{n}\right)\right) \\
&= \frac{\beta}{n-1} \left(\frac{i}{n}F^{-1}\left(\frac{i+1}{n+1}\right) + \left(1 - \frac{i}{n}\right)F^{-1}\left(\frac{i}{n+1}\right) \right. \\
&\quad \left. - F^{-1}\left(\frac{i}{n} \cdot \frac{i+1}{n+1} + \left(1 - \frac{i}{n}\right) \cdot \frac{i}{n+1}\right)\right) \\
&\leq 0 = RHS
\end{aligned} \tag{A.8}$$

(A.8) 的最后一步不等号由 $F''(x) = f'(x) \geq 0$ 在 $[0, \frac{1}{2}]$ 上恒成立的假设得出。综上, 命题 2 得证。

附录 B CFPS 实证分析结果补充

B.1 家庭教育支出的分析结果补充

除对数家庭教育支出 ($\ln eduexp$) 外, 教育支出占家庭纯收入的比重 (记为 $expind1$)、教育支出占家庭人均纯收入的比重 (记为 $expind2$) 以及教育支出占家庭储蓄的比重 (记为 $expind3$) 也是可以用于反映高中生家庭在教育上投入的指标。

$$expind1 = \frac{eduexp}{faminc} \quad (B.1)$$

$$expind2 = \frac{eduexp}{faminc_{per}} \quad (B.2)$$

$$expind3 = \frac{eduexp}{famsav} \quad (B.3)$$

按照 (B.1)、(B.2) 和 (B.3) 定义 $expind1$ 、 $expind2$ 和 $expind3$ 后, 对省份 ($province$) 和年份 ($year$) 进行二因素独立测量方差分析, 结果如图 B-1、图 B-2 和图 B-3 所示。可以发现, 结果仍均不显著。

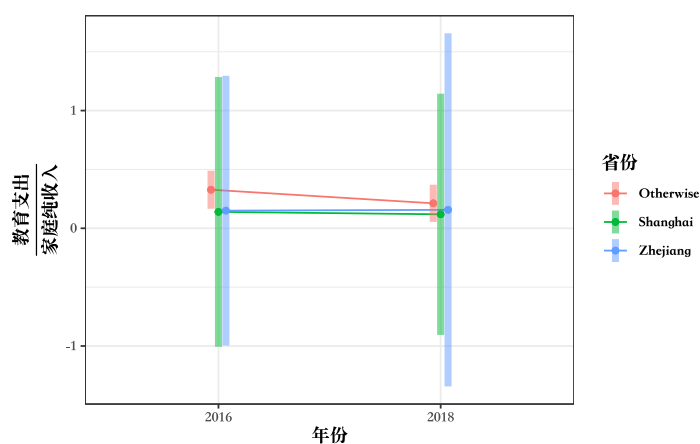


图 B-1 2016 到 2018 年教育支出占家庭纯收入比重变化趋势图 ($p = 0.985$)

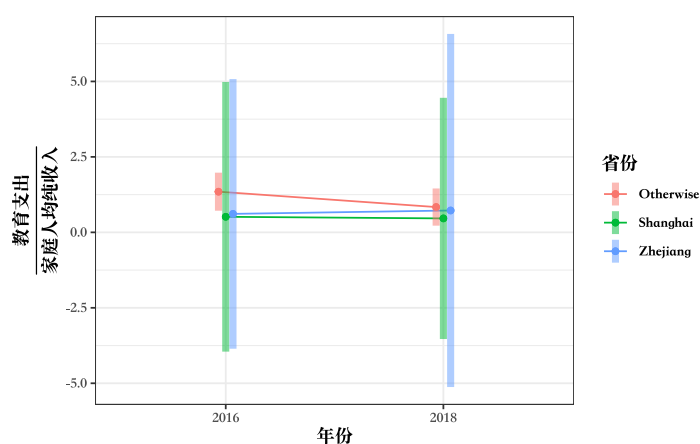


图 B-2 2016 到 2018 年教育支出占家庭人均纯收入比重变化趋势图 ($p = 0.976$)

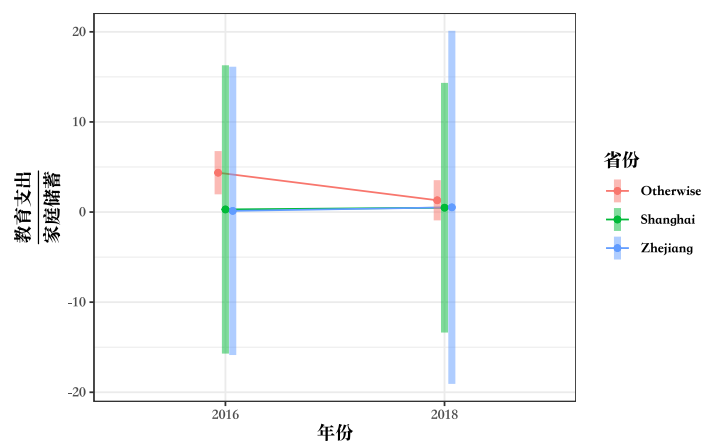


图 B-3 2016 到 2018 年教育支出占家庭总储蓄比重变化趋势图 ($p = 0.925$)

B.2 情绪指标分析结果补充

除学业压力 (*pressure*) 外, 情绪低落指标 (*depression*) 和生活愉快指标 (*happiness*) 也是能反映高中生情绪状况的指标。将二者对省份 (*province*) 和年份 (*year*) 进行二因素独立测量方差分析, 结果如图 B-4 和图 B-5 所示。可以看出, 结果仍均不显著。

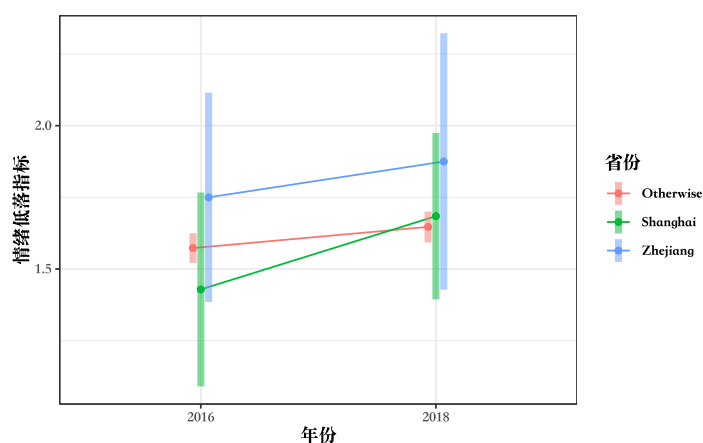


图 B-4 2016 到 2018 年情绪低落指标变化趋势图 ($p = 0.655$)

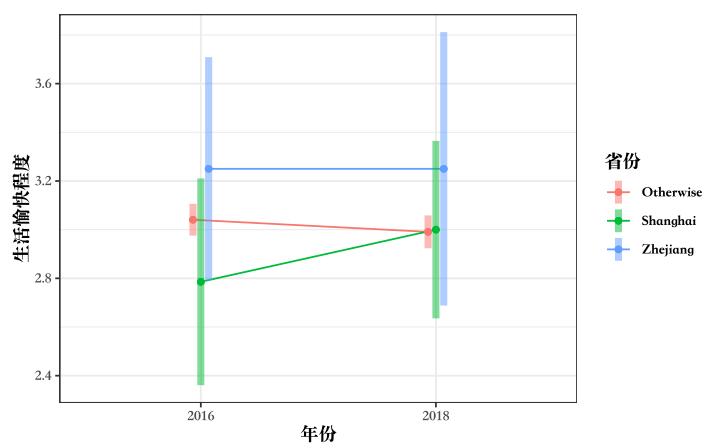


图 B-5 2016 到 2018 年生活愉快指标变化趋势图 ($p = 0.724$)

附录 C 问卷分析结果补充

C.1 问卷各变量频率分布直方图

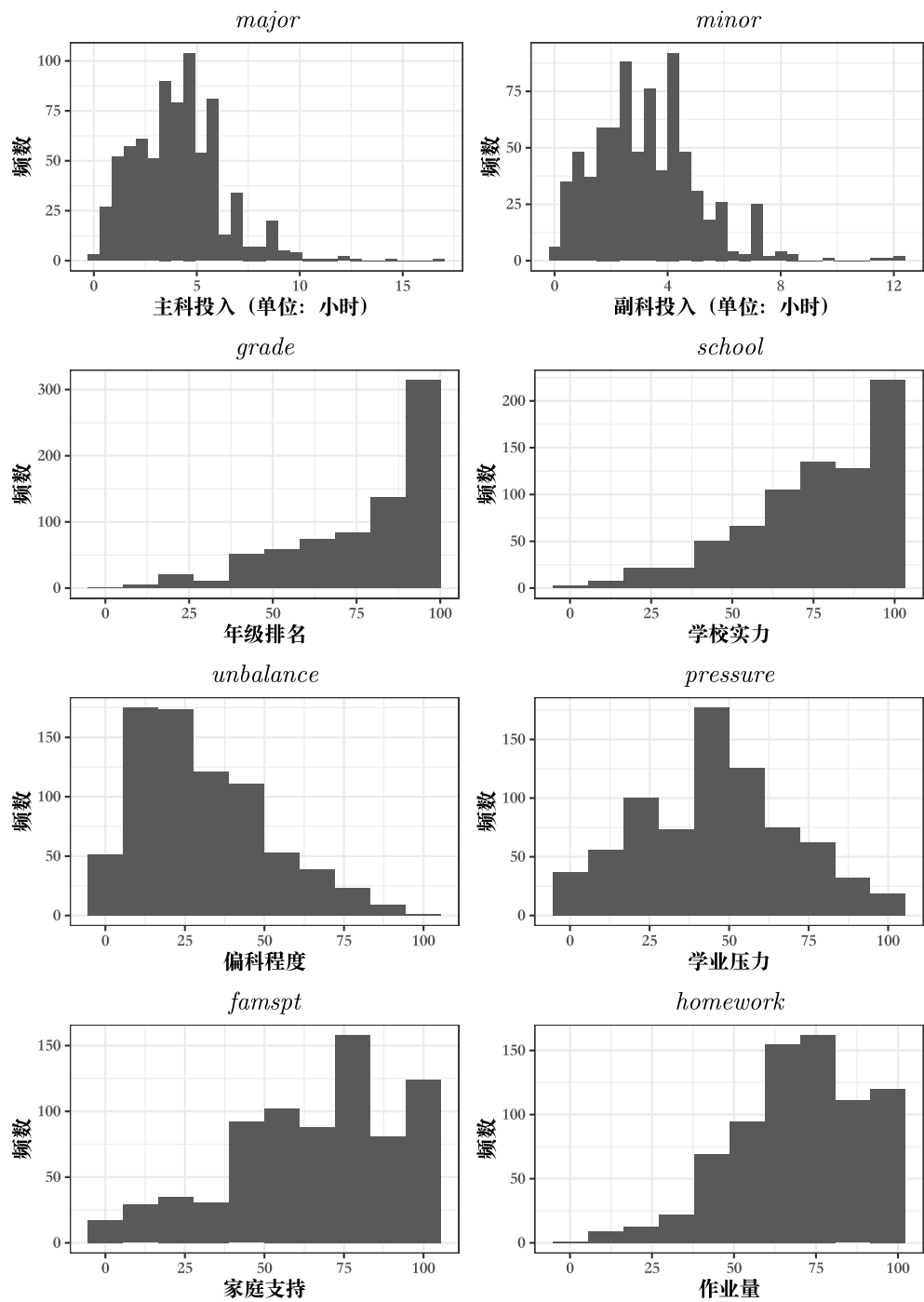


图 C-1 各变量频率分布直方图

C.2 有效答卷筛选条件

问卷经过六层筛选。第一个条件是高考年份、省份能够与其选择的高考模式相对应；第二个条件是填写的主科和选科的时间分配比之和为 100；第三个条件是问卷填写者的主科时间分配比、选科时间分配比、年级排名、学校实力、全省排名、作业量、考试频次、学习时

间均不应为 0；第四个条件是不能有成片的缺失值；第五个条件是样本不来自海南；第六个条件是高考年份不在 2015、2016、2024、2025 年。

这六层筛选的条件由强至弱。前四个条件与受访者的填写态度有关，后两个条件与研究需要有关。对于第五个条件，现行的赋分制度一共有三种，分别是一分一档、三分一档和标准分，其中海南作为高考改革的试验地，标准分目前只在海南施行；而考虑到海南施行标准分赋分的样本量很小，因此我们将海南的样本舍弃；对于第六个条件，问卷填写者参加高考的年份的分布如图 C-2 所示。

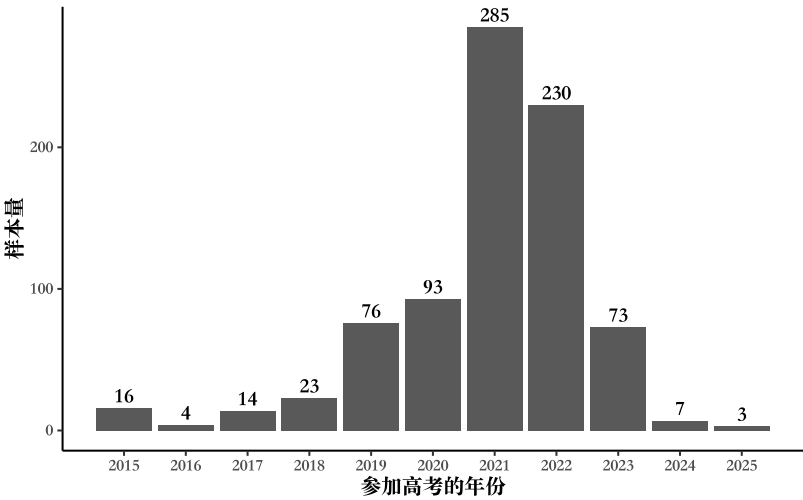


图 C-2 问卷调查样本年份分布（筛选前）

从样本分布中可以发现，2015、2016、2017、2018、2024、2025 年的考生的样本较少，由于年份是需要控制的变量，因此这部分样本将被舍弃。

回收的初始数据共有 1196 条，不通过第一个条件的样本有 278 个，不满足第二个条件的样本有 26 个，不满足第三个条件的样本有 59 个，不满足第四个条件的样本有 7 个，不满足第五个条件的样本有 2 个，不满足第六个条件的样本有 67 个。

C.3 主科和选科时间投入的模型分析结果补充

研究发现，实施三分一档改革的省份主要是北京、上海、浙江、天津等经济较发达、教育资源较丰富的省份。北京、上海、浙江和天津四地从 2019 年至 2023 年恰好覆盖了原始分、一分一档和三分一档三种政策。为了排除二者相关性对模型结果有效程度的影响，本部分仅使用北京、上海、浙江和天津四个省级行政区的样本，重做 (7.5)(7.6)(7.7)(7.8) 四个模型的回归分析。

重做后的多元线性回归（OLS）结果如表 C-1 所示。

变量	<i>major</i>	<i>minor</i>	<i>major</i>	<i>minor</i>
	(7.5)	(7.6)	(7.7)	(7.8)
<i>one</i>	-0.1265 (0.4293)	-0.9922*** (0.3729)	-0.5206 (0.5806)	-1.7070*** (0.5064)
<i>three</i>	0.0373 (0.3664)	-0.5571* (0.3183)	-0.2830 (0.4523)	-1.0975*** (0.3945)
截距项	4.1024*** (0.3469)	3.7359*** (0.3014)	-219.2317 (258.2829)	-428.0223* (225.2458)
控制变量	否	否	是	是
R^2	0.0009	0.0184	0.0678	0.0768
样本量	392	392	392	392

表 C-1 区分高考改革种类的基准回归结果（限制样本范围）

回归结果表明，无论是否控制变量，是否高考改革和赋分方式均对考生在主科上的备考时间投入无显著影响。而一分一档和三分一档相比改革前高考均能显著降低考生在选科上的备考时间投入，且前者降低更多。这一结果与不限制样本省份时的结果完全一致，说明正文中的回归结果及相关结论稳健性较高。

附录 D 问卷具体设计

D.1 说明

本问卷基于问卷星^①平台设计。
本问卷中大量使用百分滑动条题型。百分滑动条是一种让作答者通过拖动一个滑动条控件选择 0 到 100 中的数值来回答相应问题的题型，一般用于报告程度或报告占比等。百分滑动条的左右两端分别有最小值提示语和最大值提示语，作为作答者估计程度时的参照。如图 D-1 所示。



图 D-1 百分滑动条示例

D.2 问卷设计

各届高三同学高考备考情况调查

欢迎您参与此问卷调查！
本问卷将会调查您一些关于您高考时选科及备考方面的问题，如您对各个选科的擅长程度、时间投入等。我们不会询问您的具体分数、实际排名等隐私问题。
填写本问卷预计只需 5 分钟的时间。

您在填写问卷时可能会大量遇到以下题型：

【百分滑动条实例】 您认为某某的程度为？ / 您某某的占比（%）为？ [百分滑动条]
您可以利用本滑条熟悉操作。
最小值提示语： 程度极低 / 0%
最大值提示语： 程度极高 / 100%
对于这样的百分滑动条，您既可以直接填写左侧的数字，也可以拖动滑条至相应位置。
在填写问卷时，您不必对所有问题给出极其精确的答案。例如，不必纠结某个比例到底是 19%、20% 还是 21%。当您想不起来一些具体的数值（如投入时间、比例分配）时，按照模糊的记忆填写一个大概的数值或将滑条拖到一个大概位置即可。当然，精确的答案是被欢迎的。

请点击【下一页】开始作答。
[分页]

1. 您参加（或即将参加）高考的省份是？ [单选题]
选项：上海、浙江、北京、天津、山东、海南、河北、辽宁、江苏、福建、湖北、湖南、广东、重庆、黑龙江、吉林、安徽、江西、广西、贵州、甘肃、山西、内蒙古、陕西、宁夏、青海、四川、云南、河南、新疆、西藏、其他（港澳台、海外等）。

2. 您参加（或即将参加）高考的年份是？ [单选题]
选项：2015 年及以前、2016 年、2017 年、2018 年、2019 年、2020 年、2021 年、2022 年、2023 年、2024 年、2025 年、2026 年及以后。

^① 问卷星网址为 <https://www.wjx.cn>。

3. (3-1) 请判断以下陈述是否正确：您面临的高考政策是**传统文理分科模式**。[单选题]

选项：正确、错误。

(3-2) 请判断以下陈述是否正确：您面临的高考政策是 **3+3 模式**（而非传统文理分科模式）。[单选题]

选项：正确、错误。

(3-3) 请判断以下陈述是否正确：您面临的高考政策是 **3+1+2 模式**（而非传统文理分科模式）。[单选题]

选项：正确、错误。

注：如果依据作答者报告的参加高考的省份和年份推断出该省份在当年已经进行新高考改革，则展示 (3-1)；如果该省份即将进行 3+3 改革但在当年尚未进行，则展示 (3-2)；如果该省份即将进行 3+1+2 改革但在当年尚未进行，则展示 (3-3)。这保证了认真回答的作答者在本题一定选择“错误”选项。

4. (4-1) 您的选科是？[单选题]

选项：文科、理科。

(4-2) 您的选科是？[多选题，选择其中 3 项]

选项：历史、地理、政治、物理、化学、生物、(技术)。

注：若作答者在 (3-1) 选择“正确”或在 (3-2) 或 (3-3) 回答“错误”，则展示 (4-1)，否则展示 (4-2)。仅当作答者在第 1 题中选择“浙江”时，(4-2) 的“技术”选项才会出现。

[分页]

5. 您高三**上学期间**平均每天投入在**高考备考**上的学习时间约为_____小时_____分钟。
[填空题]

包括上课、完成作业、额外练习等。

如果您还没上高三，请填写您现在在课内相关学习（不包括兴趣学习）上花费的时间。

6. 您高三**放假期间**平均每天投入在**高考备考**上的学习时间约为_____小时_____分钟。
[填空题]

包括周末、寒假、国庆假等任何学校不安排课的日子中完成假期作业、额外练习等的时间。

如果您还没上高三，请填写您现在在课内相关学习（不包括兴趣学习）上花费的时间。

7. 您的上述时间投入中：

花费在**主科**（语数外）上的时间占比约为_____%

花费在**选科**（选科）上的时间占比约为_____%。[填空题]

请确保您填写的两个比例之和为 100%。

以下问题将分别询问您六门应考科目的**擅长程度**及**时间投入**。

对于擅长程度，0 为十分不擅长，100 为十分擅长。请保证您衡量各个科目擅长程度使用的自评标准基本一致。例如，如果您认为自己对语文的擅长程度为 X，对数学的擅长程度为 X+1，则您应当更擅长数学。

对于备考时间的投入占比，请回答占总备考时间的比例。例如，如果您每天只备考 1 分钟，其中 15 秒用于学习语文，则语文的备考时间投入占比应为 25%。请保证您填写的六科备考时间投入比例之和为 100%。

8. 关于**语文** [填空题，输入 0 到 100 的数字]

您认为您在多大程度上擅长语文？_____

您在语文科目上的备考时间投入占比（%）为？_____

9. 关于**数学** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长数学? _____
您在数学科目上的备考时间投入占比(%)为? _____
10. 关于**外语** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长外语? _____
您在外语科目上的备考时间投入占比(%)为? _____
11. 关于**历史** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长历史? _____
您在历史科目上的备考时间投入占比(%)为? _____
12. 关于**地理** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长地理? _____
您在地理科目上的备考时间投入占比(%)为? _____
13. 关于**政治** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长政治? _____
您在政治科目上的备考时间投入占比(%)为? _____
14. 关于**物理** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长物理? _____
您在物理科目上的备考时间投入占比(%)为? _____
15. 关于**化学** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长化学? _____
您在化学科目上的备考时间投入占比(%)为? _____
16. 关于**生物** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长生物? _____
您在生物科目上的备考时间投入占比(%)为? _____
17. 关于**技术** [填空题, 输入 0 到 100 的数字]
您认为您在多大程度上擅长技术? _____
您在技术科目上的备考时间投入占比(%)为? _____

注: 11 至 17 题中, 仅当作答者在报告的选科包含该科目时, 该科目对应的题目才会出现。

18. 您是否认为自己有明显偏科(即最强和最弱的科目差距过大)? [单选题]

选项: 是、否。

[分页]

19. 您每天平均的**睡眠时长**(含午睡)约为_____小时_____分钟。[填空题]
20. 您利用**碎片时间学习**(如在排队时背古文、在公交上背单词等)的频率为? [百分滑动条]
最小值提示语: 从不利用碎片时间学习
最大值提示语: 碎片时间总是在学习

21. 您认为您的**家庭**在您高三一年在支持您备考上的**投入**（如饭菜改善、课外辅导开支、陪伴等）程度为？[百分滑动条]

如果您还没上高三，请按照现状填写

最小值提示语：几乎没有投入

最大值提示语：几乎投入了一切

22. 您认为高三一年您的学校布置的**作业量**如何？[百分滑动条]

如果您还没上高三，请按照现状填写。

最小值提示语：几乎不布置作业

最大值提示语：几乎布置了全部可能的作业

23. 高三一年，您的学校总共组织了_____次**大型模拟考试**。[填空题]

无论是周考月考、期中期末、一模二模，只要是在短暂的几天内连续完成全部高考科目的测试并计分、排名的考试都计入。

如果您还没上高三，请按照您经历的上一个完整学年的情况回答。

[分页]

24. 您认为自己在高三时的**实力强于**多大比例（%）的**同校**高三学生？[百分滑动条]

如果您还没上高三，请按照现状回答。

最小值提示语：自己是年级里最弱的

最大值提示语：自己是年级里最强的

25. 您认为自己在高三时的**实力强于**多大比例（%）的**同省**高三学生？[百分滑动条]

如果您还没上高三，请按照现状回答。

最小值提示语：自己是全省最弱的

最大值提示语：自己是全省最强的

26. 您认为您的学校在**高考方面**的**综合实力强于**全省多大比例（%）的学校？[百分滑动条]

最小值提示语：学校是全省高考最弱的

最大值提示语：学校是全省高考最强的

27. 您在高三一年的**学习压力与焦虑**程度为？[百分滑动条]

如果您还没上高三，请按照现状回答。

最小值提示语：毫无压力

最大值提示语：压力已无法承受

28. 您认为您在高三一年的**学习压力与焦虑**程度超过了全省多大比例（%）的高三同学？[百分滑动条]

如果您还没上高三，请按照现状回答。

最小值提示语：自己是全省压力最小的

最大值提示语：自己是全省压力最大的

29. 综合考量您面临的高考难度、竞争考生人数、考生综合实力、大学名额数等因素，您认为您的**升学压力**超过了**全国**多大比例（%）的同龄人？[百分滑动条]

如果您还没上高三，请按照现状回答。

最小值提示语：自己是全国升学压力最小的

最大值提示语：自己是全国升学压力最大的

附录 E 向实验参与者发放的指导文件

实验说明与指导

欢迎大家参与此次实验！

我们正在调查大家的基本工作技能水平，本次检测包括打字和口算两项任务。您的任务是测试自己的打字速度和口算能力，并提交给我们。实验数据和结果仅用于课题组此次科学研究，若有任何顾虑，请联系我们的组织者。如没有其他问题，请继续阅读。

为了激励大家拿出自己的真实水平，我们将为两项技能“综合排名”前 2% – 10% 的人额外发放20元/人，为前 2% 的人额外发放100元/人。

对两个项目，您都可以反复尝试来提高成绩，以反映你的最高水平。界面适应是一个过程，因此我们鼓励大家多多尝试，记录最好状态。但请务必提交真实数据和历史记录哦～

[不赋分组]

您的综合成绩将按照如下方式计算：

$$\text{综合成绩} = \text{最快打字速度 (cpm)} + \text{最高口算得分 (points)}$$

提问：若一个人的最快打字速度为 20 cpm，最高口算得分为 15 points，则他的综合成绩为？

（请认真作答此题，最终提交的问卷中里的第一题即需要您填写上述题目答案）

[密赋分组]

您的综合成绩将按照如下方式计算：

我们会将所有人的最快打字速度进行排序，前 5%、5% – 15%、15% – 25%、……、85% – 95%、后 5% 将分别获得 A+、A、A-、B+、B、B-、C+、C、C-、D、E 的评定，分别对应 20、18、……、2、0 分。这个分数加上您的最高口算得分即为您的综合成绩。详情见下表。

在人群中的排序情况	等级评定	该项得分
前 5%	A+	20
5% – 15%	A	18
15% – 25%	A-	16
25% – 35%	B+	14
35% – 45%	B	12
45% – 55%	B-	10
55% – 65%	C+	8
65% – 75%	C	6
75% – 85%	C-	4
85% – 95%	D	2
后 5%	E	0

提问：若一个人的最快打字速度为 20 cpm，最高口算得分为 21 points；已知他的最快打字速度在所有人中排序为 30%；则他的综合成绩为？

（请认真作答此题，最终提交的问卷中里的第一题即需要您填写上述题目答案）

[疏赋分组]

您的综合成绩将按照如下方式计算：

我们会将所有人的最快打字速度进行排序，前 1/8、1/8 – 3/8、3/8 – 5/8、5/8 – 7/8、后 1/8 将分别获得 A、B、C、D、E 的评定，分别对应 20、15、10、5、0 分。这个分数加上您的最高口算得分即为您的综合成绩。详情见下表。

在人群中的排序情况	等级评定	该项得分
前 1/8	A	20
1/8 – 3/8	B	15
3/8 – 5/8	C	10
5/8 – 7/8	D	5
后 1/8	E	0

提问：若一个人的最快打字速度为 25 cpm，最高口算得分为 20 points；已知他的最快打字速度在所有人中排序为 2/8（即 1/4）；则他的综合成绩为？

（请认真作答此题，最终提交的问卷中里的第一题即需要您填写上述题目答案）

接下来请根据要求进入以下网站。

P.S. 若在实验中遇到技术或操作问题，欢迎联系组织者。

一、打字测试

打字测试请使用 <https://dazi.91xjr.com>（请放心进入），进入后您将看到以下界面……（后略）^①

二、口算测试

口算测试请使用 <https://www.preplounge.com/en/mental-math/add/1>（请放心进入），进入后您将看到以下界面……（后略）^②

确认结束实验后，请不要忘记**按要求填写问卷**哦。

^① 本段是提供给参与者的打字测试网站的使用说明，图片较多，篇幅较长；考虑到本部分内容并不重要，予以略去。

^② 本段是提供给参与者的口算测试网站的使用说明，图片较多，篇幅较长；考虑到本部分内容并不重要，予以略去。

Effects of Ranking-Based Categorized Scoring in National College Entrance Examination (Gaokao) on Involution of Highschool Students

CHE Tianyi CHU Zihe KAN Ziyang WANG Tianxiao
YAN Penghuan ZHOU Rui*

March 2023

Abstract: This study focuses on exploring the influence of the ranking-based categorized scoring on the degree of involution among high school students in the context of the reform of China's College Entrance Examination (Gaokao). A theoretical model is constructed based on the tournament theory and the large contest theory, and it is predicted that within the major subjects, ranking-based categorized scoring will not change the degree of involution, but it *will* affect the degree of involution within the elective subjects; under certain conditions, an increase in the density of categories in ranking-based categorized scoring will reduce the degree of involution; the influence of implementing the ranking-based categorized scoring on involution depends on specific parameters, but the results of simulations using real-world parameters show that the implementation of the ranking-based categorized scoring in reality will reduce the overall degree of involution among students. This study then uses variables such as education expenditure and pressure level of students in the CFPS database to explore the influence of the College Entrance Examination reform on involution among highschool students in Shanghai and Zhejiang. Furthermore, a questionnaire survey is conducted to obtain updated and more comprehensive samples covering more provinces, while including controlled variables in a more detailed manner. Finally, online experiments is conducted, where typing and mental arithmetic tasks are used to simulate the College Entrance Examination scenario. The experiments deeply explore the relationship between whether and of what category density ranking-based categorized scoring is adopted and the degree of involution, ultimately verifying the theoretical predictions.

Keywords: College Entrance Examination reform, ranking-based categorized scoring, involution

* All six authors are recognized as co-first authors and are ordered by family name. All authors contributed equally.