



北大教育经济研究（电子季刊）  
Economics of Education Research (Peking University)  
北京大学教育经济研究所主办  
Institute of Economics of Education, Peking University

第 19 卷  
第 1 期  
(总第 70 期)

主编：闵维方；副主编：丁小浩 岳昌君；

编辑：陈恩茹

## 分数膨胀的影响因素与学生选课策略

赵颖 哈巍

**摘要：**学生获得的分数在某种程度上反映了高校的教学质量和教学成果，同时也对高校的人才培养质量起到关键作用。本文以某研究型大学为例，探讨了分数膨胀是否发生以及发生的原因。研究表明，职称越低的教师越具有放松给分的倾向。这也许可以说明，在现有的管理和评价制度下，分数膨胀本身存在制度性的内在动机。具有竞争压力的课程分数增长最为明显，这说明教师为了吸引学生而打高分是分数膨胀不易被察觉的另一动因。此外，学生有向高分课程汇聚的倾向，并可通过改变选课策略来提高分数（组合分数膨胀），其选课策略对分数膨胀的贡献总额超过 5%。

**关键词：**分数膨胀；选课策略；双重差分；固定效应模型

作者简介：

赵颖，安徽铜陵人，北京大学政策法规研究室战略规划办公室副主任，助理研究员，研究方向为高等教育管理、院校研究；

哈巍，辽宁沈阳人，北京大学教育经济研究所研究员，北京大学教育学院副院长、长聘副教授，研究方向为教育政策评估。

### 一、引言

2001 年，《波士顿环球报》报道了哈佛学生分数持续增长的现象，2001 届毕业生获得“毕业荣誉”的比例高达 91%，引起了社会各界的广泛关注。<sup>[1]</sup>事实上，学生分数的持续增长现象并非个例。2002 年，布朗大学报告了其关于“分数贬值”问题的大型研究。2004 年，普林斯顿院长针对分数增长的问题，劝说教授们执行“评分指南”，以限制 A 等分数所占的比重。美国学者对于较长时段的分数变化进行了深入和细致的分析，梳理了 1963-2013 年全美的四年制高校分数，以及 1940-2013 年全美四年制高校 A-F 评分的比例变化，发现全美的平均 GPA 以每十年 0.1 的速度增长，A 的比例从 1940 年的 15% 上升到 2013 年的 45%，而 C、

D、F 的比例一直在下降。<sup>[2]</sup>英国也发现了分数膨胀现象，自 20 世纪 90 年代以来，英国高校的最好评分（相当于 A 等级）比例自 1996 年的不到 60% 上升到了 2012 年 72%。<sup>[3]</sup>并且，这种现象似乎已经遍布全球——加拿大、法国、瑞典等国的学者也相继观察到本国大学中存在分数的持续增长，这种现象被称为“分数膨胀”或“分数贬值”。<sup>[4]</sup>

分数是学业评价最主要的呈现形式。学业评价一般是指教师对学生学业成就的评价，是对学生某门课程成绩的综合评定与价值判断。<sup>[5]</sup>也有学者认为，学业评价是建立在事实判断基础上的价值判断。<sup>[6]</sup>然而，价值判断具有非客观的因素，因此亦有学者认为基于价值判断的评价具有“有限理性”的特征。<sup>[7]</sup>在全世界高校的大部分课程中，学业评价最后会通过教育测量的手段以量化评分的形式呈现，即为学生获得的分数。不可直接观测的学生学业成就，被坍塌到几个可观测的维度，通过可观测变量加以标定，或者说通过在不同观测变量上的观测值进行推断，因此，学业评价实际上是一种理论建构。<sup>[8]</sup>即使是同一门课程，不同教师设定的教学目标、教学内容、教学难度、测量方式等都可能存在不同，从而学生得到的学业评价也不会完全相同。尽管学业评价是建构的，无法统一教师的评价标准，但是应当具有规范性评价的内涵，即以长期以来形成的教育价值认识以及以这种价值认识为基础的教育准则为判断评价对象的标准。<sup>[9]</sup>因此，对分数膨胀现象的批判和担忧，实则是对大学培养质量和学术标准下滑的担忧。更有学者认为，这种现象是“学生消费主义”，大学里的风气发生了变化，学生们不再被认为是寻找知识的助手，相反，他们成了大学的“顾客”——教授们倾向讨好学生，给学生想要的分数，而不是他们应得的分数。<sup>[10]</sup>

分数膨胀最开始出现是带有一些积极效应的，如分数整体提高后，使更多的学生有资格申请助学贷款、获得奖学金和学术荣誉等。<sup>[11]</sup>然而，分数膨胀不可避免地产生了一些负向效应。分数膨胀实际造成了“分数贬值”。分数膨胀使得分数的分布向高分段集中，导致最好的学生和次好的学生混同在一起，无法区分。分数的集体增长，造成分数的内在价值在降低。分数膨胀也影响了研究生院招生和雇主招聘，因为分数已经不足以成为衡量一个人能力的信号。<sup>[12]</sup>同时，分数膨胀降低了学生不及格的可能，让学生认为获得学业成就不需要投入太多努力，从而损害了教学效果、影响了学风。<sup>[13]</sup>国外学者对分数膨胀现象已有较多分析和梳理。有学者考察了威廉姆斯学院（Williams College）的分数膨胀情况，发现分数发生了明显的增长并向高分段集中，其院系差异十分显著；给分高的院系课程对于学生有较高的吸引力，进而影响了学生的课程安排或专业选择。<sup>[14]</sup>有研究表明分数膨胀与教师的身份特征（终身教职教师、非终身教职教师和兼职教师）有关，认为兼职教师的聘用助长了分数膨胀的发生。<sup>[15]</sup>也有学者认为学生评教制度的实行是分数膨胀的主要成因。<sup>[16]</sup>

在经历高等教育大规模发展之后,我国已经进入提高人才培养质量的关键时期,对于影响人才培养质量的潜在因素的梳理和分析尤为重要。我国学者对分数膨胀现象亦有所关注。何山和丁小浩<sup>[17]</sup>认为分数膨胀是高等院校对就业市场压力做出的一种消极反应,他们以某大学为样本,按照时间序列,分析了该大学1992-2003年间整体平均成绩随时间的变化,并得出该大学没有发生分数膨胀的结论。陈志权等<sup>[18]</sup>将分数膨胀归因于就业市场的压力、社会诚信机制不健全、教师职业道德缺失、学生评教制度等,并认为要抑制分数膨胀,学校管理部门应当完善教学管理制度、加强教师职业道德修养、提高严格执行学业标准的自觉性。黄桂<sup>[19]</sup>利用某高校的调查和访谈问卷,探寻了分数与学生评教发生双重膨胀的原因,认为学生评教的简单粗放式运用鼓励了分数膨胀的发生。孙鳌<sup>[20]</sup>、赵颖<sup>[21]</sup>等从博弈论的角度论述了在学生评教机制下,教师利益与学生发生捆绑,教师会选择降低学术标准、提高分数来“讨好”学生以换取好的学生评教分数。笔者此前利用某大学课程优秀率放松政策作为准实验,发现政策后课程分数发生了显著上升,进而讨论了学生评教制度下教师主动放松给分而学生给予教师更慷慨的评教分数回报的这一现象。<sup>[22]</sup>比之前文,本文将给予高校分数膨胀现象更加细致的刻画,全面地分析分数膨胀背后的贡献因素,并利用某高校开放跨院系选课的政策变化作为自然实验,从学生视角进一步探讨分数膨胀的内因。

## 二、对分数增长和分数膨胀概念的厘清

分数增长是指学生学业分数发生了上涨,虽然有些文章中并未对分数增长和分数膨胀作出明确界定,但严格来说,分数膨胀应有别于分数增长,是指学生学业分数的增长超出了其实际能力的增加。<sup>[23]</sup>这里对分数增长和分数膨胀的概念进行进一步的厘清。

分数的持续增长看似只是教师的给分发生了变化,但事实上分数增长并不是一个孤立事件,它是众多因素综合作用的结果。根据以往的研究,主要有以下几种:

**其一,高等教育大众化下的高校扩招。**在高等教育大众化的进程中,高校扩招顺应了社会经济发展的需求,让越来越多的年轻人有机会进入大学学习。这产生了两个问题:一是高校之间的生源争夺,二是高校面临着前所未有的来自劳动力市场的压力,为了增强毕业生在劳动力市场上的竞争能力,教师在给分时较以往更宽松。<sup>[24]</sup>更为关键的是,原本精英化的高等教育在走向大众化的过程中,高校育人的理念在发生变化,学生人数上升使得大学的门槛越来越低,原先的“精英标准”在松懈,取而代之的是较低的学术要求。

**其二,教师自身的发展诉求。**在现行的评教体制下,配合高校人事制度的改革(预聘—长聘制),评教分数影响着教师的晋升、评奖和评优,因此教师往往会通过放松给分来提高

自己的学生评教分数，由此可能引发分数膨胀。<sup>[25]</sup>此外，给学生较高的分数也有利于师生和谐相处，避免师生关系的对立与紧张。并且，教师也能减少因学生挂科、无法毕业等事务所带来的时间上的占用、心理上的压力和“不必要”的麻烦，减轻教学负担，从而把更多经历投入到科研上。

**其三，学生对于分数的功利主义倾向。**竞争的压力使得学生越来越在乎分数，学生会主动向教师索要高分，这也给教师带来了心理上的压力，迫使部分教师放松给分。研究表明，学生索分要分的潜在威胁会导致教授放松给分标准。<sup>[26]</sup>并且，对于高分的追逐，使得学生也倾向于选择容易课、送分课，来提高自己的学分绩。

**其四，学生素质的提高。**学生本身的素质提高了，有了更好的学习能力，从而获得了更高的分数。有学者试图用学生 SAT 成绩的提高来说明学生入学质量的提高<sup>[27]</sup>，然而也有不同的意见认为 SAT 并不是一个很好的 GPA 预测指标。<sup>[28]</sup>由于我国高考分数在不同省份、不同年份间不可比，因此对于学生素质的变化趋势较难衡量，但也许可以通过弗林效应来佐证学生素质的提高是一种必然。弗林发现了发达国家智商测验结果呈现逐年增长的现象，被称为弗林效应。<sup>[29]</sup>2012 年，弗林将这一结果扩展到某些发展中国家，也发现了类似的效应。<sup>[30]</sup>根据弗林效应的观测结果，每十年，美国、日本和中国智商测验得分平均增长了 3 分、7.7 分和 4.5 分。弗林效应形成的原因，主要有两个方面，一是环境因素，如更好的教育环境、更多复杂信息的刺激、更小的家庭规模等；另一个是生物学因素，如营养状况的提高、卫生医疗条件的改善等。其中，更好的教育环境被认为是最为关键的因素。<sup>[31]</sup>而这些因素，也同样可以促使学生素质不断提高。通过弗林效应，或许有理由相信，学生素质的持续提高是非常合理的现象，而且应当对分数增长有所贡献，但这一效应的显现是较为长期的。

从成因来看，前三类因素导致教师降低给分标准，引发分数膨胀，而学生素质的提高带来的分数增长是积极影响，不能归纳于分数膨胀之中。因此，在研究分数膨胀时，应当从分数增长中剥离学生素质的变化带来的影响。如有学者利用哈佛大学本科生入学考试成绩来测量学生学术水平的初始程度，认为学生大学阶段分数的提高与学生入学考试成绩的提高密切相关。<sup>[32]</sup>也有学者使用克莱姆森大学学生学术能力评估测试（SAT）数据，认为分数增幅的 25% 归因于学生素质的提高。<sup>[33]</sup>

### 三、背景和数据

本研究使用国内某研究型大学（下简称 A 大学）2012-2017 学年度共 10 个学期（以下表述为第 1-10 学期）的本科生课程平均成绩数据，所属学科涵盖理科、工科、人文、社科、医学等 5 个大类。因体育课和理论课的教学过程有所区别，因此去掉全部体育课程，只保留理论课程，共得到 15994 条数据。

由于我国高考试题的难度存在年份差异，因此不同年份的生源质量难以比较。在 5 年的观察期内，该校生源素质保持稳定，暂且视为学生素质的变化对分数增长的贡献应当较小，分数增长的部分即为分数膨胀。另外，在此观察期内，该校课程评分方式未发生明显变化。

需要说明的是，本文使用的数据皆为课程层次数据，因此，在所有涉及平均分之处，都应使用课程人数做加权，具体公式如下：

$$G_i = \frac{\sum_{j=1}^n (g_{ij} Enrollment_{ij})}{\sum_{j=1}^n Enrollment_{ij}}$$

其中， $g_{ij}$  是某门课程平均分（算术平均数）， $Enrollment_{ij}$  是某门课程上课人数， $G_i$  是加权平均分，数据描述如表 1 所示。

表 1 各类变量描述性统计

变量	(A)2012-2017	(B)2012-2013	(C)2013-2014	(D)2014-2015	(E)2015-2016	(F)2016-2017	(G)Change(F-B)
优秀率 <sup>①</sup>	0.51 (0.23)	0.48 (0.22)	0.50 (0.23)	0.51 (0.23)	0.53(0.23)	0.55(0.24)	0.07
平均分	83.51 (5.35)	83.08 (5.15)	83.36 (5.17)	83.49 (5.12)	83.53(5.55)	84.03(5.65)	0.95
平均学分	2.40 (1.22)	2.47 (1.17)	2.40 (1.27)	2.40(1.22)	2.38(1.22)	2.37(1.19)	-0.10
平均上课人数	48.05 (57.90)	52.51 (60.84)	50.03 (58.61)	47.73(57.47)	45.37(55.97)	45.35(56.62)	-7.15
课程数	15994	2997	3010	3220	3349	3418	421

①该大学将 85 分及以上视为优秀，某门课达到优秀的人数占比即为该门课的优秀率。

从表 1 来看，2012-2017 年的平均优秀率为 51%，2016-2017 学年相对 2012-2013 学年优秀率整体提高了 7%，且 A 大学整体优秀率一直在上升。2012-2017 年的课程平均分为 83. 51，2016-2017 学年相对 2012-2013 学年平均分上升了 0. 95 分。5 年来平均学分下降了 0. 1。从平均上课人数看，5 年来一直在减少，2016-2017 学年比 12-13 学年平均每门课上课人数减

少了 7.2 人，即相对减少了 14%。上课人数的减少是因为课程数量在增加，2012-2017 这 5 年平均每年增加 84.2 门课。

A 大学的课程类型基本分为全校性必修课、通选课和专业必修课、专业选修课。在 2016 年春季学期（第 8 学期）之前，除了全校性必修课和通选课之外，A 大学的学生不能直接跨院系选修其它院系的课程，只能选择本院系的专业课程。自 2016 年秋季学期（第 9 学期）起，A 大学开放了跨院系选课，并鼓励学生选择其它院系的课程。如果学生的选课行为或偏好发生了较大的变化，比如说向高分课程汇聚，那么也可能是造成分数增长的原因。跨院系选课的开放给同学们的选课行为变化带来了可能，也使得我们得以观察选课行为变化带来的分数增长。

课程平均分随学期的变化趋势见图 1。由图 1 可以初步看出，全校课程平均分一直在增长，由第 1 学期的 82.9 分上升到第 10 学期的 84.2 分，平均每学期增长 0.14 分。

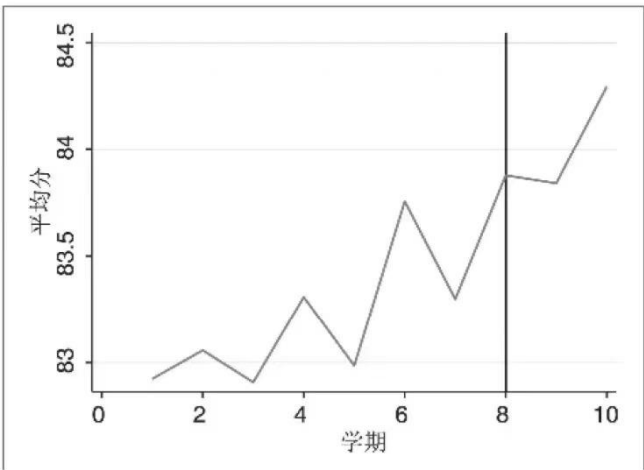


图 1 平均分随学期的变化趋势

此外，由图 2 可以看出，整体而言各学科课程平均分都有上升的趋势。STEM 学科课程的总体平均分和上涨幅度均最小，从第 1 学期的 82.3 分上升到第 10 学期的 83.4，增长了 1.1 分，平均每学期增长 0.12 分。人文类课程上涨幅度最大，5 年平均分增长了 2.6 分，平均每学期上涨 0.26 分。经管类课程和社科类课程平均分分别上涨了 1.7 分和 1.4 分。并且，在放开跨院系选课，非 STEM 学科课程有了更明显的上涨趋势，而 STEM 课程的平均分反倒

有所下降。进一步把 STEM 课程和非 STEM 课程做两样本 T 检验，可知 STEM 课程比非 STEM 课程的平均分显著低 1.2 分。

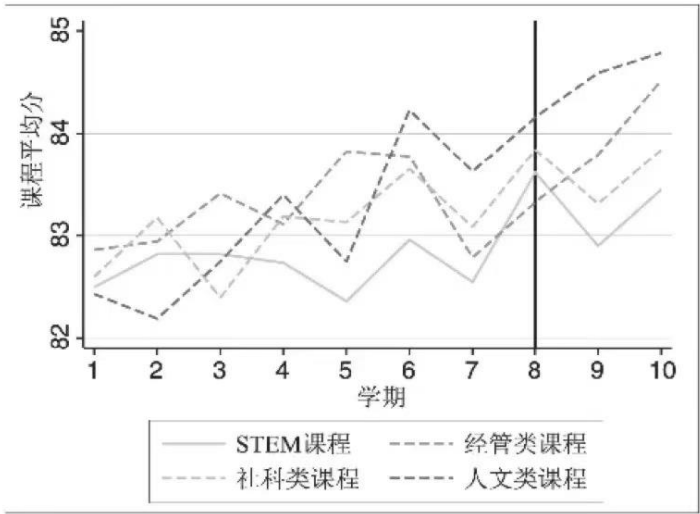


图 2 不同学科课程平均分随学期的变化趋势

图 3 按照课程性质，把全部课程分为了专业课（包括专业必修和专业选修）以及除专业课以外的非专业课（包括全校必修、通选课等）。可以看到，1-10 学期的专业课平均分呈一定的上升趋势，从第 1 学期的 82.9 分上升到第 10 学期的 83.9 分，总增长 1 分。而非专业课程平均分增长的更为明显，从第 1 学期的 83.2 分上升到低 10 学期的 85.2 分，总增长达到 2 分之多。为确定院系之间的组间差异和院系内的组内差异，以院系为单位做方差分析，结果见表 2。由方差分析可知，以院系为单位分组，组间差异存在显著的差异，且组间差异约是组内差异的 45 倍。可知各院系的给分有很大的不同。

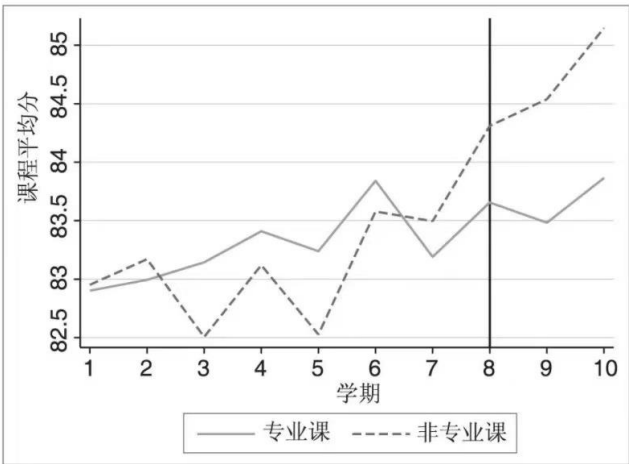


图 3 专业课和非专业课平均分随学期的变化趋势

表 2 方差分析结果

	SS	df	MS	f	Prob>F
组间差异	45985.15	40	1149.63	45.73	0.0000
组内差异	460266.24	18309	25.14		
总计	506251.39	18349	27.59		

#### 四、方法和模型

从描述统计可以看到,A校课程5个学年的平均分提高了约1分,增长了1.13%,略高于全美GPA的增长速率(全美平均GPA以每十年0.1的速度增长,5年的增长幅度约1.0%)。因此,有必要进一步考察分数的增长是否显著,以及推动分数增长的具体因素。

##### (一) 课程及教师背景因素对分数的影响

为了考察分数的相关影响因素,且进一步推断分数增长的具体因素,引入课程和教师的相关变量,建立固定效应模型如下:

$$G_{ijt} = \varphi_0 + \alpha_t + \beta Course_i + \gamma Ins_j + \sum_{k=1}^k \varphi_k x_{ijtk} + \epsilon_{ijt}$$

其中, $\alpha_t$ 是学期固定效应, $Course_i$ 是课程固定效应, $Ins_j$ 是教师固定效应, $x_{ijtk}$ 是一系列控制变量,包括课程相关控制变量,如课程学分数、选课人数、课程性质,以及教师相关控制变量,如教师职称、教师年龄、教师教龄(自来校时间算起)、教师学历、教师性别。

因使用控制变量时可能存在遗漏变量的问题,因此进一步使用固定效应。为避免共线性造成的不显著,当加入教师固定效应时,应去除教师相关控制变量;当加入课程固定效应时,应去除课程相关控制变量。

##### (二) 学生的选课策略——高分课程对学生的“虹吸”效应

学生选课行为的变化,如向高分课程聚拢,也是造成分数膨胀的诱因之一。因A大学选课方式的改革,使得我们可以检验分数增长是否是由于学生选课行为发生变化造成的。在2012-2017年这5年中,我们需要检验学生们的选课行为是否发生了变化,以及这些变化对于分数变化的贡献。



为了检验高分课程对学生是否会产生“虹吸”效应，使得学生越来越向高分课程汇聚，本文将借用 A 大学选课方式的改革作为准实验进行因果推断，将政策改变前平均分大于等于 85 分的课程作为实验组，低于 85 分的课程作为对照组。政策改变后，由于学生可选范围大大增加了，在选择其它院系课程时，有可能会将往期课程给分高低作为选课的重要依据之一。因此，在政策改变后，我们可能会观察到，相对于对照组，实验组的上课人数在增加。特别的，我们将分析的个体定义为同一位老师教授的同一门课程。建立双重差分（DID）双向固定效应模型如下：

$$E_{ij} = \beta_0 + \beta_1 Treated_i + \beta_2 Postpolicy_j + \beta_3 Treated_i * Postpolicy_j + \gamma_i + \theta_j + e_{ij}$$

其中， $E_{ij}$  代表某个体  $i$ （教师和课程联合变量） $j$  学期课程的上课人数。 $Treated_i$  是一个虚拟变量，当它等于 1 时该个体属于实验组，等于 0 时该个体属于对照组。 $Postpolicy_j$  是一个虚拟变量，当它等于 1 时表示政策改变后，等于 0 时表示政策改变前。 $Treated_i * Postpolicy_j$  是一个交互项，其系数  $\beta_3$  即为本文关注的处理效应。 $\gamma_i$  代表个体（教师和课程联合变量）固定效应，它控制的是不随时间变化的课程性质或教师特质。 $\theta_j$  是学期固定效应，控制的是不随个体（老师和课程）而变化的总体时间趋势，如该学校不同年级学生群体在学业能力上的差异，该校整个教师或者学生群体对于教学或者学习的重视程度的总体变化趋势等。当这两个固定效应同时使用时， $Treated$  和  $Postpolicy$  这两个变量将被吸收掉，因此没有必要再展示其系数。 $e_{ij}$  是随机误差项。

## 五、回归结果

### （一）课程及教师背景因素对分数的影响

回归结果见表 3。在表 5 中，第（1）个模型只放入学期控制变量，以测量课程平均分随学期变化的趋势并作为基准。可以看出，自第 4 学期开始，分数发生了显著的增长。

表 3 课程分数的相关影响因素

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
课程学分		-0.456*** (0.0527)	-0.490*** (0.0557)	-0.484*** (0.0558)			
选课人数		-0.0185*** (0.000839)	-0.0165*** (0.000920)	-0.0164*** (0.000919)			
课程性质（以通选课为参照组）							
专业必修		-1.890*** (0.120)	-1.770*** (0.135)	-1.773*** (0.134)			
专业选修		-0.864*** (0.123)	-0.756*** (0.141)	-0.762*** (0.139)			
全校必修		-3.942*** (0.208)	-4.071*** (0.223)	-4.074*** (0.222)			
双学位		0.187 (0.580)	-0.509 (0.703)	-0.508 (0.709)			
教师职称（以讲师为参照组）							
教授			-0.370** (0.186)	-0.287 (0.187)			
副教授			-0.104 (0.153)	-0.0645 (0.153)			
教师年龄			0.00833 (0.00953)	0.00430 (0.00960)			
教师教龄			-0.0243*** (0.00828)	-0.0261*** (0.00826)			
教师学历 (是否博士)			-0.0209 (0.136)	-0.0687 (0.137)			
教师性别 (是否男性)			-0.671*** (0.0979)	-0.668*** (0.0979)			
学期固定效应（以第1学期为参照组）							
第2学期		0.213 (0.189)		0.233 (0.188)	-0.0490 (0.182)	-0.177 (0.189)	-0.140 (0.208)
第3学期		0.229 (0.186)		0.240 (0.197)	0.0129 (0.153)	-0.0396 (0.146)	-0.155 (0.154)
第4学期		0.459** (0.185)		0.231 (0.195)	0.110 (0.172)	-0.0414 (0.180)	-0.132 (0.204)
第5学期		0.355** (0.179)		0.329* (0.186)	0.0135 (0.149)	0.0713 (0.141)	-0.0477 (0.147)
第6学期		0.596*** (0.185)		0.314 (0.192)	0.253 (0.170)	0.155 (0.195)	0.162 (0.204)
第7学期		0.304* (0.176)		0.416** (0.184)	0.0252 (0.153)	0.0676 (0.148)	-0.109 (0.158)
第8学期		0.760*** (0.198)		0.567*** (0.200)	0.329* (0.174)	0.128 (0.201)	0.0703 (0.227)
第9学期		0.673*** (0.205)		0.690*** (0.217)	0.458*** (0.173)	0.407** (0.166)	0.350** (0.169)
第10学期		0.918*** (0.202)		0.806*** (0.209)	0.599*** (0.180)	0.365* (0.192)	0.234 (0.213)
教师固定效应					是	是	是
课程固定效应							
常数项	82.98*** (0.128)	86.67*** (0.134)	87.10*** (0.371)	86.90*** (0.386)	83.22*** (0.116)	83.34*** (0.118)	83.41*** (0.133)
N	15,994	15,994	15,994	15,994	15,994	15,994	15,994
R <sup>2</sup>	0.002	0.084	0.088	0.090	0.445	0.566	0.711

注：（1）\*\*\* $p<0.01$ ，\*\* $p<0.05$ ，\* $p<0.1$ ；（2）括号内为稳健标准误

表 5 学生选课行为的变化及其对分数增长的贡献

院系名称	平均上课人数 (5年)	2017年相比 2012年上课 人数变化	2017年相比 2012年课程 平均分变化	相对平均分 (和全校 均值比较， 5年) (A)	上课人数占全校的占比				对平均分变 化的贡献 (A*B)
					2012-2017年	2012年	2017年	占比变化 (B)	
上课人数占比减少的院系									
A1	62.03	-25.47	-0.58	-0.82	3.6%	4.66%	3.10%	-1.56%	0.0128
A2	59.58	31.18	3.15	-0.8	1.7%	1.86%	0.60%	-1.26%	0.0101
A3	33.93	-4.39	1.16	-0.82	2.7%	3.41%	2.38%	-1.03%	0.0084
A4	29.01	0.14	3.8	-2.46	7.1%	6.98%	6.19%	-0.80%	0.0196
A5	50.45	-14.07	0.16	0.76	3.2%	3.66%	2.89%	-0.77%	-0.0058
.....									
上课人数占比增加的院系									
B1	46.15	-3.83	2.07	0.76	3.2%	2.94%	3.36%	0.42%	0.0032
B2	64.68	-10.69	3.06	2.04	3.8%	4.14%	4.56%	0.42%	0.0085
B3	49.44	3.29	0.4	-0.39	3.4%	2.88%	3.33%	0.45%	-0.0017
B4	78.6	-17.53	1.8	-2.17	3.8%	3.40%	3.93%	0.54%	-0.0116
B5	50.47	4.81	0.46	-1.12	4.7%	4.23%	4.92%	0.69%	-0.0077
.....									
总变化									0.0494

第（2）个模型加入了课程相关变量，包括课程学分、上课人数和课程性质。可以看出，课程学分每增加 1 个，课程平均分减少 0.46 分；选课人数每增加 1 个，课程平均分减少 0.02 分。总体而言，通选课给分最高，比专业必修课平均高出 1.9 分，比专业选修课平均高出 0.9 分，比全校必修课高出 3.9 分，比双学位课程无显著差异。全校必修课几乎是所有学生必须要上的课，包括军事理论、计算机基础、马克思主义课程等，且大多以固定分班的形式，学生没有太多自选的余地，因此也没有吸引学生选课的压力。专业必修课也有类似于全校必修课的性质，然而，在目前逐渐开放学生选课和转专业的形势下，专业必修课要承担起部分吸

引外系学生、巩固本系学生的任务，因此，需要有一定的吸引学生的功能。专业选修课一般面向本系或本专业同学，因为非必选课程，所以有一定的竞争压力。而双学位课程实质为学生单独付费的课程，且需要吸引外系学生来上，因此，也具有竞争的压力。基本可以看出，课程给分的高低和课程性质背后所固有的竞争压力存在相关性，竞争越充分的课程，教师愿意给出的分数越高。

第（3）个模型在第（2）个模型的基础上加入了教师相关变量，包括教师职称、教师年龄、教龄、学历和性别。讲师比教授平均给分显著高出 0.37 分，和副教授相比并不显著。教师年龄对给分的影响并不显著，而教龄（此处为入校教龄）越高的教师越有严格给分的倾向，教龄每高 1 年，平均给分减少 0.02，换言之，新入职的教师更倾向于给高分。学历似乎也 and 给分高低无关。男性教师相对女性教师给分平均低 0.67。

第（4）个模型在第（3）个模型的基础上加入了学期固定效应。第（5）个模型同时固定了学期和教师。第（6）个模型同时固定了学期和课程。第（7）个模型同时固定了学期、课程和教师。在固定了教师效应后，各学期的系数减小了，10 个学期的分数总增长降为 0.60；在固定了课程效应后，各个学期的系数减小得更多，大部分学期都变得不显著，10 个学期的分数总增长降低到 0.37。可见，课程对分数增长的贡献比教师的贡献要大，除了教师自身的特征在影响给分高低，课程的学分、性质等固有属性对分数的影响更大。教师在打分时，不仅考虑到给分对课程评教分数的影响，更会考虑这门课程的授课内容、课程性质、面向学生群体等课程相关特性来决定给分的高低。当同时固定教师和课程时，10 个学期的分数总增长降低到 0.23，且不显著。换言之，教师自身的特征和课程本身的特征[34]是引发分数膨胀的主要因素，而和学生相关的因素，如学生素质的提高、学生的性别和来源等，对分数膨胀的影响较小、贡献率较低。

根据表 5 的数据，将调整后的分数曲线作图，如图 4 所示。在学期固定效应的基础上加入教师固定效应后，分数的增长幅度从 0.81 降低到 0.60，这说明，对分数膨胀的贡献中，有 0.21 分来自于老师自身的背景特质；而更多的影响来自于课程——当固定课程效应之后，分数的增长幅度降低到 0.37，进一步下降了 0.23 分。也就是说，老师因为职称晋升等需要有放松给分的动力，而越来越大的课程竞争压力，也构成了分数膨胀的另一个不易被察觉的因素，这一因素贡献了 24% 的影响。进一步同时控制课程和教师后，分数增长进一步下降了 0.13，这可以认为是教师的背景特质和课程的不同性质发生作用时，对于课程给分的影响。而其它的因素包括学生素质的提高等，约贡献了 30% 的影响。

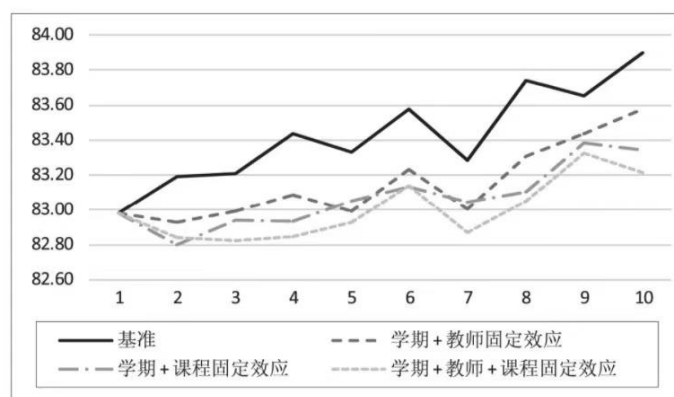


图 4 调整后的分数曲线

## (二) 学生的选课策略——高分课程对学生的“虹吸”效应

由表 4 可以看出,开放选课政策后,相对于对照组,实验组的平均上课人数多出 3.2 人。同时,因为全校必修和通选并不存在跨院系选课的问题,因此,在政策改变后,这两种性质的课程并没有发生显著的变化,而政策主要影响的专业课程都发生了不同程度的效应。相对于对照组,给分高的专业必修课平均上课人数增加了 5 人,给分高的专业选修课平均上课人数增加了 3 人。可见,当学生的选课范围扩大后,学生也不会盲目选课,或只根据兴趣选课,而是同样会考虑到该门课程的给分情况。高分课程对学生的“虹吸”效应确然存在。

表 4 双重差分回归结果

	(1)	(2)	(3)	(4)	(5)
	总体	专业必修	专业选修	全校必修	通选
处理效应	3.218***	5.016***	3.124***	2.326	-2.518
	(0.950)	(1.532)	(1.118)	(3.345)	(3.788)
常数	51.07***	56.05***	28.54***	48.72***	81.91***
	(1.020)	(1.815)	(1.186)	(2.626)	(5.076)
个体固定效应	是	是	是	是	是
学期固定效应	是	是	是	是	是
课程门数	15,183	5,977	5,454	494	1,866
$R^2$	0.927	0.917	0.879	0.929	0.946

注: 1.括号中为稳健标准误; 2. \*\*\* $p<0.01$ , \*\* $p<0.05$ , \* $p<0.1$ ; 3.课程总门数有所不同, 是因为只在政策后出现的课程没有参与回归, 因此去掉了这些课程

表 5 检验了学生跨院系选课行为的变化及这种变化对于分数的变动造成的影响(此处仅展示了部分院系的数据)。与 2012 年相比, 2017 年各院系平均每门课的学生上课人数发生了明显的变化。有些院系, 如 A2 院系, 平均每门课上课人数增加了 31 人; 而有些院系, 如 A1 院系, 平均每门课上课人数减少了 25 人。而平均上课人数的变化有可能是因为开课数

量的改变，因此，这并不能成为衡量学生选课行为变化的指标。这里选择使用某院系上课的总人数在全校上课人数的占比作为衡量学生选课行为的指标，并用这一指标在 2012-2017 年间的变化来衡量学生选择行为的变化。2012 年到 2017 年，所有院系该占比的变化在-1.56%~1.42%之间，变化的幅度并不大。

由表 5 我们还可以计算出 2012-2017 年间各院系或单位对平均分变化的贡献。以 A1 学院为例，该院系相对全校平均分均值少 0.82 分，而 5 年间上课人数占比全校变化了-1.56%，则该院系对 5 年间平均分变化的贡献为  $(-0.82) * (-1.56\%) = 0.0128$ 。其它院系类同，可以计算出全部院校或单位对平均分变化的贡献总额为 0.05 分，占 5.3%（平均分 5 年间共变化 0.95），这即是学生跨院系选课对分数增长的贡献。

## 六、总结与讨论

在高等教育内部，对科学研究活动进行量化管理在全世界都有上升的趋势[35]，这引起了诸多讨论和反思。学业评价一直具有量化的传统，其定量管理的属性反倒容易被忽视。当学生的学业成就被简化成几个易测量可比较的指标，教师在外界的压力下就具有评价标准偏移的倾向，学生为了自身的利益，也可能追逐好的评价结果而较为忽视自身的能力获得。

本文通过对 A 大学的实证研究，得到了一些初步的观察结论。A 大学生源稳定，在 5 年内不存在生源明显的下降或优化，因此，本文认为，5 年内该校学生的素质基本持平，学生素质的提高这一原因的贡献应当较小。然而，A 校的课程平均分数发生了较为明显的上涨，超过了全美增长的速率，且从回归分析的结果来看，这种上涨是显著的。从教师背景的角度分析，职称越低的教师越具有放松给分的倾向。这可能是因为，职称较低的教师面临更大的晋升压力，一方面，他们可能会选择给学生高分以避免不必要的麻烦，从而把更多的时间留给科研；另一方面，他们也更在意和学生建立良好的互动，以在学生评教中获得更好的分数或在学生中建立良好的口碑。[36]从课程属性的角度分析，本文发现，课程固定效应对分数增长的贡献比教师特征的贡献更为明显，这表露出分数增长的一个不易被察觉的动因——当引入新教师时，新开课的数量有了可观的增加，平均每门课的上课人数一直在减少。

A 大学跨院系选课政策的变化，使本文得以观察学生的选课策略。政策变化后，高分专业课发生了明显的“虹吸”效应，学生向高分课程汇聚。在以院系为分析单位的基础上，本文对学生跨院系选课的变化进行了刻画和分析，研究结果表明，院系间的选课行为变化对分

数膨胀的贡献率占 5.3%——并非由教师放松给分，而是由学生选课策略的变化造成的分数膨胀被成为“组合分数膨胀”[37]，这种提法在以往的文献中较少涉及。在政策变化前，学生可能已然倾向于选择给分高的课程，但由于这种选课策略在院系内部进行，容易形成“代代相传”的惯习沉淀下来，从而难以观察。政策后，学生仍然可以在院系内部选择高分课程。因此，学生选课策略对分数膨胀的贡献应超过 5.3%。并且，由于课程规模的限制，当高分课程被“争抢一空”后，剩下的学生只能转选给分较低的课程，所以，假设取消课程规模的限制，高分课程的“虹吸”效应会更加明显。学生对于课程的“理性”选择，值得深思，这可能代表，对于分数的追求、过于量化的评价方式限制了学生更多元的选择和更多样的发展。

值得说明的是，课程和教师背景因素对分数膨胀的影响与学生选课策略对分数膨胀的影响是不同的。教师背景和课程属性基本决定了一门课的给分情况，而往期给分情况又影响了学生选课，所以实际上，学生选课带来的分数膨胀部分，在固定了教师效应和课程效应后，也被吸收掉了。换言之，课程和教师背景因素通过影响分数，进而也影响了学生选课。笔者之前的研究里，重点讨论了学生评教制度下教师为了迎合学生，主动给学生放松给分的现象。教师之所以迎合学生，可能就是因为观察到，学生向高分课程集中、给高分课程老师以更好的评教分数回报等现象。而教师的“放水”，给了学生更多寻找高分课程机会，让越来越多的学生裹挟进来，从而使得分数膨胀愈演愈烈。本研究从学生的角度进一步考察了分数膨胀和教师主动放松给分相结合，更能看出，教师和学生结成了所谓的“共谋”。学生对于高分的追逐，既是分数膨胀的原因，又是分数膨胀的结果。

本文的分析基本印证了以往文献中分数膨胀的原因，除课程竞争的压力这一较为隐晦的因素之外，基于教师职称晋升的压力、学生对高分的追求，分数膨胀的发生十分自然。我国已有高校关注到这一现象的发生。如清华大学在《建立促进学生全面发展的学业评价体系改革方案》中提到：“从对近 20 年来清华学生成绩的统计分析来看，清华存在明显的‘分数膨胀’现象”[38]，并且，在 2019 年 4 月发布的《清华大学关于完善学术评价制度的若干意见》中明确指出：“学生评价要有利于树立优良学风，有利于促进学生全面发展”。

纽曼在《大学的理想》中谈到：“大学教育有非常实际、真实、充分的目的，不过，这一目的不能与知识本身相分离。知识本身即为目的。”[39]然而实际上，教师会根据课程的属性和自身的需要选择不同程度的放松给分，此时，虽然知识传递仍然是重要的，但不可避免地发生了退让——让学生更满意、有利于学生利益、使自己和学生更能适应社会需要即实现了眼前短期的教育目的。更令人担忧的是，学生对于知识、能力和德性的追求似乎已经让位于“分数”，他们追求数字上的“好看”，甚至把分数当做自己最大的利益。学生作为学

习者，似乎已经默认接受教育的目的是更好地适应社会经济需求，用文凭和成绩充当信号，以保证自己能够“交换”到良好的升学和就业机会为第一诉求。也许，如何让教师回归教学本位，让学生回归学习本真，以及为二者创造回归的条件，是目前亟需解决的重要问题。

本文仍有一些局限留待将来解决。首先，由于缺乏学生高考成绩的数据，以及各年份高考难度不可比等问题，本文无法剥离学生素质的提高这一因素，只能假设在 5 年这一时段里，学生素质没有发生大的变化，从而认为在观察期内分数增长即为分数膨胀。其次，本文的数据时段较短，缺乏较长时段的观察。最后，本文的研究结论基于一所高校的情况，缺乏对我国高校整体情况推广的有效依据。

## 参考文献:

- [1]哈瑞·刘易斯.失去灵魂的卓越[M].侯定凯译.上海:华东师范大学出版社,2006.92-93.
- [2]Stuar Rojstaczer and Christopher Healy, "Grading in American Colleges and Universities," Teachers College Record (2010): 1-6.
- [3]Ray Bachan, "Grade Inflation in UK Higher Education," Studies in Higher Education 42, no.8 (2017): 1580-1600.
- [4]Faieza Chowdhury, "Grade Inflation: Causes, Consequences and Cure," Journal of Education and Learning 7, no.6 (2018): 86.
- [5]张学立,谢治菊.大学生学业评价体系改革:动因、问题及对策[J].中国大学教学,2017,(6): 89-96.
- [6]刘尧.论教育评价的科学性与科学化问题[J].教育研究,2001,(6): 22-26;原文为“教育评价”,因“学业评价”包含在“教育评价”之中,为突出本文所谈对象,此处将该文原意改为“学业评价”。
- [7]Ya-Wen How et al., "Student Evaluation of Teaching as a Disciplinary Mechanism: A Foucauldian Analysis," The Review of Higher Education 40, no.3 (2017): 325-352.
- [8]杨向东.教育测量在教育评价中的角色[J].全球教育展望,2007,(11): 15-25.
- [9]刘志军.教育评价的反思和建构[J].教育研究,2004,(2): 59-64.
- [10]Stuar Rojstaczer and Christopher Healy, "Grading in American Colleges and Universities," Teachers College Record (2010): 1-6.
- [11]John A.Pearce, "How Employers Can Stanch the Hemorrhaging of Collegiate GPA Credibility," Business Horizons 60, no.1 (2017): 35-43.
- [12]Wan-Ju Iris Franz, "Grade Inflation under the Threat of Students' Nuisance: Theory and Evidence," Economics of Education Review 29, no.3 (June 2010): 411-22.
- [13]John.A.Pearce, "How Employers Can Stanch the Hemorrhaging of Collegiate GPA Credibility," Business Horizons 60, no.1 (2017): 35-43.
- [14]Richard Sabot and John Wakeman-Linn, "Grade Inflation and Course Choice," The Journal of Economic Perspectives 5, no.1 (1991): 159-170.
- [15]Boualem Kezim et al., "Is Grade Inflation Related to Faculty Status?" Journal of Education for Business 80, no.6 (2005): 358-364.
- [16]Kiridaran Kanagaretnam et al., "An Economic Analysis of the Use of Student Evaluations: Implications for Universities," Managerial and Decision Economics 24, no.1 (2003): 1-13; Tim Ehlers and Robert Schwager, "Honest Grading, Grade Inflation, and Reputation," CESifo Economic Studies 62, no.3 (2016): 506-521.
- [17]何山,丁小浩.从对分数膨胀现象的分析看高等院校应对劳动力市场的压力[J].清华大学教育研究,2004,(6): 35-42.



- [18]陈志权等.大学“分数膨胀”现象探析[J].化工高等教育,2015,(3):104-108.
- [19]黄桂.分数膨胀与等级膨胀:评教系统双重失效原因探析——基于某部属重点高校大学生评教的视角[J].高教探索,2011,(6):95-103+127.
- [20]孙鳌.分数膨胀的博弈分析[J].现代大学教育,2016,(5):23-27.
- [21]赵颖.学生评教制度下分数膨胀的内在逻辑:博弈模型的建立[J].中国高教研究,2019,(4):20-26.
- [22]哈巍,赵颖.教学相“涨”:高校学生成绩和评教分数双重膨胀研究[J].社会学研究,2019,(1):84-105.
- [23]Faieza Chowdhury, “Grade Inflation: Causes, Consequences and Cure,” *Journal of Education and Learning* 7, no.6 (2018): 86.
- [24]Martin O.Grady and Simon Quinn, “The Causes of Grade Inflation: an Exploration.Social and Institutional Pressures and Policy Choices,” [https://www.researchgate.net/publication/242264618\\_The\\_Causes\\_of\\_Grade\\_Inflation\\_An\\_Exploration\\_of\\_Social\\_and\\_Institutional\\_Pressures\\_and\\_Policy\\_Choices](https://www.researchgate.net/publication/242264618_The_Causes_of_Grade_Inflation_An_Exploration_of_Social_and_Institutional_Pressures_and_Policy_Choices).
- [25]Jan Tucker and Bari Courts, “Grade Inflation in the College Classroom,” *Foresight* 12, no.1 (2010): 45-53.
- [26]John A.Pearce, “How Employers can Stanch the Hemorrhaging of Collegiate GPA Credibility,” *Business Horizons* 60, no.1 (2017): 35-43.
- [27]Rey Hernández-Julián and Adam Looney, “Measuring Inflation in Grades: An Application of Price Indexing to Undergraduate Grades,” *Economics of Education Review* 55, (2016): 220-232.
- [28]S.Rojstaczer, “Grade Inflation at American Colleges and Universities,” [www.gradeinflation.com](http://www.gradeinflation.com).
- [29]王爱平等.中国儿童智力测验的弗林效应及其影响因素[J].测量与评价,2016,(5):3-11.
- [30]Edward Dutton et al., “A Flynn Effect in Kuwait, 1985 - 1998,” *Personality and Individual Differences* 138, (2019): 355-357.
- [31]Robert L.Williams, “Overview of the Flynn Effect,” *Intelligence* 41, (2013): 753-764.
- [32]哈瑞·刘易斯.失去灵魂的卓越[M].侯定凯译.上海:华东师范大学出版社,2006.119.
- [33]Jan Tucker and Bari Courts, “Grade Inflation in the College Classroom,” *Foresight* 12, no.1 (2010): 45-53.
- [34]因教师和相关课程的特征亦是影响学生是否选课的关键因素,所以由学生选课策略变化带来的分数膨胀,绝大部分也被这两个因素吸收掉了。
- [35]Don Brenneis et al., “Getting the Measure of Academia: Universities and the Politics of Accountability,” *Anthropology in Action* 12 (2005): 1-10.

[36]何山,丁小浩.从对分数膨胀现象的分析看高等院校应对劳动力市场的压力[J].清华大学教育研究,2004,(6):35-42.

[37]Talia Bar et al., "Grade Information and Grade Inflation: The Cornell Experiment," The Journal of Economic Perspectives, no.3 (2009): 93-108.

[38]清华大学.建立促进学生全面发展的学业评价体系改革方案[Z].2015.

[49]约翰·亨利·纽曼.大学的理想[M].徐辉等译.杭州:浙江教育出版社,2001.