

我国“双一流”大学 平均学分绩点 (GPA) 制度调查与反思

——基于内容分析法

曹文振 王涛利

摘要: 在“双一流”建设目标指引下,我国高等教育迈向国际化,源于欧美国家的平均学分绩点 (GPA) 成为大学生学业评价的新制度并出现了“分数膨胀”(Grade Inflation)现象,但强调微观层面个体利己主义的过往研究视角或许掩盖了背后的结构性制度困境。研究采用内容分析法对“双一流”建设高校成绩评价制度的关键要素开展编码与分析,发现“双一流”高校的学生成绩评价制度存在明显差异,GPA生产过程有着清晰的制度线索,“分数膨胀”不仅是个体行为的结果,更是制度逻辑的产物。基于此,研究提出GPA转换体系应有合理敏感度,突出其作为分级评价与原始分不同作用;建立具有可比性的GPA制度框架,增强其在跨校评价中的作用;优化GPA评价结果的理性分布,兼顾教师享有的评价自主权;改进GPA评价制度的整体公平性,为学习困难学生留下“容错”空间。

关键词: 平均学分绩点; GPA; “双一流”建设高校; 成绩评价制度; 大学生; 学业表现

一、研究缘起

改革开放以后,我国高校对大学生学业成就的评价制度经历了从不完备到逐渐完善,乃至体系化的过程。自2005年教育部颁布《普通高等学校学生管理规定》提出“考核和成绩评定方式,由学校规定”^[1]以后,政府对高校的控制权开始下放,高校对学生考核和评定的自主权逐渐扩大,学分制改革的进程不断加快。在2017年教育部正式提出建设“世界一流大学和一流学科”的目标以后,中国大学成绩评价体系逐渐与欧美高校并轨,作为舶来品的平均学分绩点 (Grade Point Average, 简称GPA) 成为愈发盛行的学习质量评价指标。在学生评奖评优、推荐免试攻读研究生、海外升学等特定情景下,原始分数发挥的作用被削弱,GPA的地位愈加突出,俨然成为了成绩评价的代名词;在就业市场中,GPA也是有吸引力的衡量依据^[2]。

在这一背景下,与西方高校类似的“分数膨胀”(Grade Inflation)现象随之而来,成为我国高等教育的新痛点。美国精英大学在上世纪就开始面临“分数膨胀”问题,在普遍实行平均学分绩点制度后,学生成绩严重向高分段集中^[3]。例如杜克大学1961年平均GPA为2.7,到1994年时已上升到3.3^[4];华盛顿大学也从1964年的2.31上升到3.12^[5]。美国艺术与科学院(AAAS)在2002年发布了《评价与学院:我们做得对吗》报告,其中将分数膨胀定义为“在一个较长时期内,学生的学业成就未提高,但其GPA却明显上升的现象”^[6]。与海外情况类似的是,我国高校学生的GPA也呈现出逐年上升的趋势,且在“双一流”高校中表现尤为明显。以天津大学为例,自2009年起短短三年学生的GPA就从2.95升至3.17^[7];2015年清华大学颁布了《“建立促进学生全面发展的学业评价体系”改革方案》,毫不避讳地指出该校存在严重的“GPA

收稿日期: 2020-05-30

作者简介: 曹文振,北京大学信息管理系博士研究生;王涛利,通讯作者,北京大学教育学院博士研究生。(北京/100871)

导向学业成就观”和“分数膨胀”现象^[8]。

许多大学的行政部门都曾试图采取措施降低“分数膨胀”，但由于对产生的原因莫衷一是，常常导致政策的失败与抵抗^{[9][10]}。已有研究多从现象本身或是理论视角出发探究这一现象产生的缘由。基于问题导向的研究发现，教师在对学进行学业评价时已不再单纯基于能力和学业表现，受到新自由主义思潮影响，学生对高分的追求被当作消费者需求得以满足^{[11][12][13]}。基于理论导向的研究则多采用“经济人”和“理性人”假设为视角，认为高校内部在各利益相关者功利主义心态膨胀的氛围中，教师与学生可能会通过学生成绩和评教分数“双向膨胀”的方式满足自身利益诉求^{[14][15]}。前人多将利己主义视作“矛盾根源”的研究取向，造成囿于个体层面的分析难以突破功利主义视角的束缚。值得反思的是，高校所设定的学生学业评价制度才是教师开展分数评定的合法性来源，但学业评价制度在这一过程中的深层影响却被当前的研究忽视，使得对现象背后因素的探究存在结构性失察。因而，本研究希望将考察视野从微观个体转移到宏观制度层面，通过深入评价制度本身追溯高校学生成绩评定的运作机制，以期挖掘造成 GPA 上涨的根源性因素，建立看待“分数膨胀”现象的不同视角，并基于新的发现提出有力的解决对策。

二、资料与方法

(一) 数据来源与总体情况

结合教育部发布的“双一流”高校建设名单，本研究从各高校官方网站中的“信息公开”等栏目、教务处等职能部门网站以及其他文件汇编材料（如学生手册）中获取与本科生成绩评价的有关政策文本。除政策文件外，各高校颁布的学分制管理办法、学分绩点算法方案、本科生成绩单等也是该研究主题的重要印证和补充材料。经过梳理与验证原始文件颁布时间，研究选取各高校最新修订的文件进行分析，共纳入 41 所“双一流”建设高校的学业评价政策文本作为分析素材，具体情况见表 1。

受教育部等发布的“双一流”建设实施办法（暂行，2017 年）的影响，2017 年成为高校对相关文件进行更新的重要时间节点。有 34 所入选“双一流”建设名单的高校都在 2017 年及以后新制定或修订了本科生成绩评价政策文本（其中 2017 年 23 所、2018 年 5 所、2019 年 6 所）。我国高校的本科生学

业成绩评价政策主要由教务处（部）颁布并同时享有对规定的解释权，这部分高校共有 36 所。其次，有 5 所高校的本科生院是学生成绩的管理部门，还有部分学校则由党委学生工作部、招生办公室、注册中心等职能部门参与有关工作。

表 1 “双一流”建设高校本科生学业成绩评价政策文本概览

政策类型	高校名称	颁布或解释权 所属部门	颁布时间
成绩 管理 规定/ 办法	中国人民大学	教务处	2015 年 3 月
	浙江大学	本科生院	2015 年 3 月
	北京师范大学	教务处	2015 年 11 月
	吉林大学	教务处	2016 年 9 月
	北京理工大学	教务处	2017 年 6 月
	华东师范大学	教务处	2017 年 6 月
	哈尔滨工业大学	教务处	2017 年 7 月
	上海交通大学	教务处	2017 年 9 月
	电子科技大学	教务处	2017 年 9 月
	大连理工大学	教务处	2018 年 8 月
成绩 评定/ 记载 办法	山东大学	本科生院	2018 年 9 月
	东南大学	教务处	2017 年 8 月（注 2）
	厦门大学	教务处	2012 年 5 月
	清华大学	注册中心、教务处、 研究生院	2015 年 5 月
学籍 管理 规定/ 办法	北京大学	教务部	2019 年 6 月
	中国海洋大学	教务处	2019 年 6 月
	中南大学	本科生院	2017 年 6 月
	同济大学	教务处、招生办公室	2017 年 7 月
	华中科技大学	教务处	2017 年 7 月
	中央民族大学	教务处	2017 年 8 月
	天津大学	教务处	2017 年 8 月
	复旦大学	教务处	2017 年 8 月
	四川大学	教务处	2017 年 8 月
	华南理工大学	教务处	2017 年 8 月
	重庆大学	教务处	2017 年 8 月
	西安交通大学	教务处	2017 年 8 月
	东北大学	教务处	2017 年 9 月
	云南大学	教务处	2017 年 9 月
	西北农林科技大学	教务处	2017 年 9 月
	新疆大学	教务处、 党委学生工作部	2017 年 9 月
	西北工业大学	教务处	2017 年 10 月
	湖南大学	教务处	2017 年 10 月
	北京航空航天大学	教务处	2017 年 12 月
	郑州大学	教务处	2017 年 12 月
	武汉大学	本科生院	2018 年 5 月
	中国科学技术大学	教务处	2018 年 7 月
	兰州大学	教务处	2018 年 12 月
	中国农业大学	本科生院	2019 年 5 月
	南开大学	教务处	2019 年 7 月
	南京大学	教务处	2019 年 8 月
	中山大学	教务处	2019 年 8 月

* 注 1：国防科技大学因学校性质特殊，2018 年才首次招收非军籍本科生，课程管理与成绩评价资料缺失；

注 2：东南大学最新学分制管理办法未标注颁布日期，最早可查时间为 2017 年 8 月。

(二) 研究方法

由于高校成绩评定制度的原始文本过长且较为繁琐，需要对其进行概念提取，并对相似概念展开重编码，研究采用内容分析法的研究范式进行编码方案的设计和提取。内容分析法是社会科学领域一种专门用以对文本内容进行分析的惯常方法，最大

的优势是能够揭示出文本中的隐含事实及内容^[16]。由于本研究没有可供借鉴的前人研究框架，研究者在广泛阅读政策文本并征求有关专家学者意见的基础上，经过多次讨论并结合我国高校成绩评定与 GPA 形成方式的特点及对部分文本的试编码情况，确定内容分析所需的编码框架方案，一级类目包括成绩评定形式、过程评价要求和重修成绩评定等七类，并细分为 28 个二级类目，具体内容如下：

表 2 学业成绩评定政策文本分析编码框架

一级编码	一级类目	二级编码	二级类目	编码单元
A	成绩评定形式	A1	考核形式	考试、考查、无
		A2	成绩记载形式	百分制、等级制、二者兼有
		A3	是否允许放弃或更改课程成绩	允许、不允许
B	过程评价要求	B1	过程评价最低占比 (%)	按实际情况编码
		B2	过程评价最高占比 (%)	按实际情况编码
C	重修成绩评定	C1	单门课程最多重修次数	按实际情况编码
		C2	是否覆盖原始成绩	覆盖、不覆盖、无法判断
		C3	是否允许重修已通过课程	允许、不允许
D	GPA 算法方案	D1	单门课程 GPA 转换方式	原始分、连续化、区间化、连续区间混合、无
		D2	单门课程 GPA 转换方法	按实际情况编码
		D3	全部课程 GPA 转换方式	连续化、无
		D4	全部课程 GPA 转换方法	按实际情况编码
		D5	非百分制成绩是否参与 GPA 计算	参与、不参与、无法判断
		D6	重修课程成绩参与 GPA 计算方式	每次成绩均计入、以首次通过成绩计入、以最高一次成绩计入、以最新一次成绩计入、无法判断
E	GPA 评价尺度	E1	档次数量	按实际情况编码
		E2	内部最多分档数量	按实际情况编码
		E3	最高绩点 (满绩点)	按实际情况编码
		E4	最低通过成绩对应绩点	按实际情况编码
		E5	未通过成绩对应绩点	按实际情况编码
		E6	绩点变化最小差距值	按实际情况编码

(续表 2)

一级编码	一级类目	二级编码	二级类目	编码单元
F	优秀档次转化方式	F1	最高成绩对应档次	A+、A、优、无
		F2	获得优秀档 (A 档) 成绩下限	按实际情况编码
		F3	是否设置无额外绩点的荣誉档次	设置、未设置、无
G	成绩分布要求	G1	是否要求呈正态分布	是、否
		G2	优秀率最高比例 (%)	按实际情况编码
		G3	优秀率最低控制分数或档次	按实际情况编码
		G4	不及格率最高比例 (%)	按实际情况编码
		G5	是否允许突破成绩分布限制	允许、不允许、无规定

根据编码框架形成的编码簿，由两名研究者对 41 所“双一流”大学的政策文本进行逐一编码。为了提高内容编码的效度，参与编码的成员首先明确了编码规则和类目内涵，而后分别独立开展互不干扰的编码工作。为了对编码信度进行检验，研究采用了 Cohen’s Kappa 系数来考察两位编码者的一致性，其具体的计算方法为：

$$K = \frac{p_o - p_e}{1 - p_e}$$

其中 p_o 代表实际编码结果中不同编码者的结论一致性比例， p_e 是理论上预期的一致性概率。一般而言，Kappa 系数值达到或超过 0.75 即可认为编码结果的一致性较为可靠。本研究的编码结果 Kappa 系数为 0.932，不同编码者的一致性检验结果较优。对于不一致的编码结果，研究者引进一位非原始编码成员，再次进行讨论，直至所有编码最终形成共识。

三、“双一流”大学学生成绩评价关键要素识别

(一) 成绩评价方式

表 3 成绩评价方式政策文本编码结果

一级编码	一级类目	二级编码	二级类目	编码单元	频数	比例 (%)
A	成绩评定形式	A1	考核形式	考试、考查	32	78.05
				无	9	21.95
		A2	成绩记载形式	百分制	6	14.63
				等级制	4	9.76
				二者兼有	31	75.61
		A3	是否允许放弃或更改课程成绩	不允许	36	87.80
				允许	5	12.20

(续表 3)

一级编码	一级类目	二级编码	二级类目	编码单元	频数	比例(%)
B	过程评价要求	B1	过程评价最低占比(%)	30	7	17.07
				40	3	7.32
				50	1	2.44
				无	30	73.17
		B2	过程评价最高占比(%)	50	2	4.88
				70	3	7.32
				无	36	87.80
C	重修成绩评定	C1	单门课程最多重修次数	不限	35	85.37
				1次	2	4.88
				2次	3	7.32
				3次	1	2.44
		C2	是否覆盖原始成绩	覆盖	11	26.83
				不覆盖	25	60.98
				无法判断	5	12.20
		C3	是否允许重修已通过课程	允许	13	31.71
				不允许	28	68.29

编码结果表明,考试与考查并行是课程考核的主流形式;除少数高校未作说明外,32所高校均采用了这种方式。百分制是“双一流”高校最为重要的成绩记载形式,有37所高校(90.24%)采用了这种形式。但受到前述的学分制和绩点制改革影响,以五级制和通过制为主的等级制正在成为另一种重要形式,超过75%的高校目前都使用百分制、等级制并存的成绩记载方式。

通常情况下,学生修读过的课程均纳入GPA计算范围,补考、重修、缓考等非正常情况下取得的成绩也会被单独标记或另行记载。但在政策文本编码过程中发现,有5所“双一流”高校(12.20%)允许学生放弃或更改不满意的课程成绩,或是通过选修其他同性课程取代原始课程成绩。例如,清华大学允许学生每学期最多放弃一门任选课成绩,华南理工大学则规定通识教育课程不合格的学生可选修其他课程进行替代。

“双一流”建设高校多采用过程性评价和终结性评价相结合的方式评定课程成绩,即平时成绩与期末成绩按一定比例折算后计入最终成绩。有11所高校(26.83%)为了加强对过程性评价的管理,还对过程性评价在最终成绩中的占比作出了明确规定,这些高校规定的平时成绩最低占比在30%–50%之间浮动,有的还规定了50%–70%的平时成绩最高比例。

对于没有通过的课程,所有“双一流”高校都

提供重修的途径让学生能够重新获得学分。少数高校限制了单门课程最多的重修次数,但大多数学校(35所,85.37%)在培养期内可以反复修读未通过的课程。针对多次修读同一门课程的情况,有的学校在出具成绩单时以重修成绩覆盖原始成绩(26.83%),有的学校则将每一次得分都予以登载(60.98%)。为了能够获取更高的学分绩点,重修已通过课程被学生称为“刷课”,但编码结果表明,真正提供“刷课”途径的“双一流”高校仅有1/3不到(13所),重修已通过课程并非政策范围内提高GPA的主流方式。

(二) GPA 算法方案

表4 GPA 算法方案政策文本编码结果

一级编码	一级类目	二级编码	二级类目	编码单元	频数	比例(%)
D	GPA 算法方案	D1	单门课程 GPA 转换方式	原始分	4	9.76
				连续化	8	19.51
				区间化	22	53.66
				连续区间混合	5	12.20
				无	2	4.88
		D2	单门课程 GPA 转换方法	共 27 种	—	—
		D3	全部课程 GPA 转换方式	连续化	39	95.12
				无	2	4.88
		D4	全部课程 GPA 转换方法	共 2 种	—	—
		D5	非百分制成绩是否参与 GPA 计算	参与	17	41.46
				不参与	18	43.90
				无法判断	6	14.63
		D6	重修课程成绩参与 GPA 计算方式	每次成绩均计入	7	17.07
				以首次通过成绩计入	5	12.20
				以最高一次成绩计入	5	12.20
				以最新一次成绩计入	9	21.95
				无法判断	15	36.59

高校需要对学生获得的原始成绩(实得成绩)进行相应的绩点转换,才能够得出GPA。除北京理工大学和西北工业大学未在成绩评价政策文本中明确GPA制度,仅以实际考试分数作为衡量标准外,其余39所“双一流”高校均在政策文本中提供了GPA转换方式及其方法说明。

与此同时,不少“双一流”高校在引进西方

GPA 制度时都进行了本土化调整,调整内容包括纳入计算的课程性质、绩点浮动范围、等级划分等等,形成了直接使用原始分 (9.76%)、连续化 (19.51%)、区间化 (53.66%)、连续区间混合 (12.20%) 等多种转换形式,具体转换方案多达 27 种。连续化转换算法得到的是按百分制成绩计算的连续曲线,例如南京大学采用“课程原始分÷20”的转换方式来计算绩点。区间化转换方式意味着每一个分数段折合成对应档次的绩点。连续区间混合制则代表在一定的分数段内,绩点呈现出连续变化的曲线 (连续制),但在达到某个临界值以后绩点即保持固定不变 (区间制),电子科技大学的绩点折算标准就是此种转换方式的代表 (X 代表百分制成绩,下同):

$$\text{课程绩点} = \begin{cases} 4.0 & (85 \leq X < 100) \\ (X - 45) / 10 & (60 \leq X < 84) \end{cases}$$

无论采用哪种形式的换算方法,课程绩点与原始成绩都呈正相关关系,差异主要体现在反映原始成绩变化的敏感度不同,比如同济大学采用了比较粗粒度的转换方式,不同档次之间的绩点差距为 1;大连理工大学采用了以 0.1 为变化区间的细粒度计算方法;北京大学和北京航空航天大学采用了完全连续化的计算公式,原始分的细微变化都会在保留小数点后两位的绩点上得以体现,其具体计算公式为:

$$\text{课程绩点} = 4 - 3(100 - X)^2 / 1600 \quad (60 \leq X < 100)$$

对于全部课程 GPA 的计算,美国高校普遍采用的计算方式是 $\Sigma(\text{原始成绩对应绩点} \times \text{课程学分}) / \Sigma \text{课程学分}$ 。38 所“双一流”建设高校均沿用了这种典型算法,仅有哈尔滨工业大学制定了特殊的 GPA 算法,将不及格的考查课学分纳入计算范畴,该校制定的具体算法如下:

$$\text{平均学分绩} = \frac{\sum \text{考试课程绩点} \times \text{课程学分} - \sum \text{考查课不及格学分数}}{\sum \text{学期数}}$$

除了明确 GPA 的计算方法以外,不少高校还对纳入计算的成绩范围进行了具体的规定。比如,对于通过制 (P/NP) 这样的非百分制成绩,参与 GPA 计算 (17 所) 与不参与 GPA 计算 (18 所) 的学校基本相当;又如对于重修课程成绩参与 GPA 计

算的方式,涵盖了每次成绩均计入 (7 所)、以首次通过成绩计入 (5 所)、以最高一次成绩计入 (5 所)、以最新一次成绩计入 (9 所) 四种形式。

(三) GPA 评价尺度

表 5 GPA 评价尺度政策文本编码结果

一级编码	一级类目	二级编码	二级类目	编码单元	频数	比例 (%)
E	GPA 评价尺度	E1	档次数量	5	18	43.90
				无	23	56.10
		E2	内部最多分档数量	1	4	9.76
				3	14	34.15
				无	23	56.10
		E3	最高绩点 (满绩点)	5	10	24.39
				4.8	1	2.44
				4.5	1	2.44
				4.3	4	9.76
				4	19	46.34
				无	6	14.63
		E4	最低通过成绩对应绩点	3	1	2.44
				2	2	4.88
				1.7	1	2.44
				1.5	4	9.76
				1.3	2	4.88
				1.2	1	2.44
				1	24	58.54
				无	6	14.63
		E5	未通过成绩对应绩点	连续	1	2.44
				0	34	82.93
				无	6	14.63
		E6	绩点最小变化尺度	连续	14	34.15
				0.1	1	2.44
				0.2	2	4.88
				0.3	16	39.02
				1	2	4.88
				无	6	14.63

在 GPA 评价制度中设置了档次的的所有 18 所“双一流”建设高校都采用了五档评价方式,即将学生成绩划分为 A、B、C、D、F 五档。其中,14 所高校在档次内部还进行了+/-档的细分,如将 A 档细分为 A+、A、A-三个内部档次。

在 GPA 的制度设计上,不同高校转换后的绩点上下域值存在差异。从“满绩”(即最高绩点)来看,“双一流”高校共规定有 5 种满绩点,最大可有 1 分的差距 (4.0-5.0)。以 4.0 为满绩的高校最多 (46.34%),其次为 5.0 满绩 (24.39%),也有部分高校采取了 4.3、4.5、4.8 等非整数形式的满

绩点。美国研究型大学主要采用满绩点为 4.0 的 GPA 作为学生学业的整体评价方式^[17]，这再次反映出我国高校在引进西方制度时融入了本土化思考。“双一流”高校在最低通过成绩（通常为 60 分或及格档次）所对应绩点上的范围差异更大，存在 2 分的浮动（1.0-3.0），但采用 1.0 作为通过课程绩点下限的学校最多（58.54%），其他则分布较为零散。对于未通过课程的成绩，几乎所有高校的 GPA 绩点方案都计为 0 分，无论百分制原始成绩位于 0-60 分中的哪个位置，仅有南京大学规定“不及格的课程在补考或重修通过之前按实得分计入学分”。

在最高与最低绩点的区间范围内，不同档次之间绩点变化的最小尺度一般为 0.3 分，或者采用不设档次的连续值（即用平滑曲线涵盖所有绩点，不划分区间段），设置这两种变化尺度的高校分别有 16 所（39.02%）和 14 所（34.15%）。

（四）优秀档（A 档）转化方式

表 6 优秀档（A 档）转化方式政策文本编码结果

一级编码	一级类目	二级编码	二级类目	编码单元	频数	比例 (%)
F	优秀档转化方式	F1	最高成绩对应档次	A+	10	24.39
				A	6	14.63
				优	1	2.44
				无	24	58.54
		F2	获得优秀档（A 档）成绩下限	85	9	21.95
				86	1	2.44
				87	1	2.44
				90	11	26.83
				无	19	46.34
		F3	是否设置无额外绩点的荣誉档次	设置	4	9.76
				未设置	14	34.15
				无	23	56.10

前人研究表明，“分数膨胀”更容易出现在精英高校追求高分档次的学生中^[18]。从编码结果可以看到，“双一流”大学对高分档次的绩点转换方案规定得尤为详细。在“优”“A”等本身代表优秀学业表现的档次之上，有 10 所高校（24.39%）还细分出“A+”档次来授予最优秀的学生；甚至有 4 所高校的“A+”仅仅是一种表彰学业突出学生的荣誉档次，获得“A”与“A+”的学生在绩点上完全相同。

当百分制原始成绩被转换为对应的等级和绩点时，“优秀”档次成为评价的醒目标签。“双一流”大学对取得优秀档次的分数下限进行了以下两种设

计：参考传统的百分制标准，将 90 分作为优秀档次分界线；为了让更多学生能够进入“优秀”档次，将对应的原始成绩放宽到 87、86 和 85 分（两种方式均有 11 所高校采用）；不存在将优秀档次对应的分数线提高的学校。当然，也有近一半的高校（46.34%）仅是将原始分数进行绩点转换，未赋予档次的标签加以区分。

基于上述两部分的分析可以发现，由于高校在百分成绩与等级成绩之间的对应转换方案不同，产生了高校间“同分不同等”或“同分不同绩”的现象，相同分数在连续制、等级制（GPA4.0、GPA5.0）标准下的绩点可能相距甚远。以百分制原始成绩 85 分为例（见表 7），不同 GPA 转换制度下的对应绩点范围从 3.3 分直至 4.5 分，有着最高 1.2 分的差距；对应的档次则可以是“良”、“B+”、“B”、“A-”。即使同为北京地区的“双一流”高校，85 分在中国人民大学里的绩点为 3.3（GPA4.0），等级为 B+；中央民族大学的绩点为 3.5（GPA5.0），等级为良；而在同样采用 GPA4.0 制度的中国农业大学里，绩点为 3.7，等级为 A-。若以 GPA 作为校际之间学生学业表现的比较标准，由此可见并不存在太多适用空间，反而容易造成诸多困惑和混淆。

表 7 不同 GPA 转换制度下原始分对应绩点与档次情况（85 分）

对应绩点（满绩点）	对应档次	高校名称
4.5 (5)	-	兰州大学
4.25 (5)	-	南京大学
4 (5)	-	浙江大学
4 (4)	-	华中科技大学、电子科技大学
3.7 (4.3)	A-	中国科学技术大学、西安交通大学、上海交通大学
3.7 (4)	A-	中国农业大学、四川大学、厦门大学、复旦大学
3.7 (4)	-	中国海洋大学、郑州大学、西北农林科技大学、武汉大学、天津大学
3.58 (4)	-	北京大学、北京航空航天大学
3.5 (5)	良	中央民族大学
3.5 (5)	A-	山东大学
3.5 (5)	-	中山大学、东北大学、大连理工大学、新疆大学
3.5 (4.8)	B	东南大学
3.5 (4.5)	-	湖南大学
3.5 (4)	-	重庆大学、云南大学
3.3 (4)	B+	中国人民大学、吉林大学

(五) 成绩分布要求

表 8 成绩分布要求政策文本编码结果

一级编码	一级类目	二级编码	二级类目	编码单元	频数	比例(%)
G	成绩分布要求	G1	是否要求呈正态分布	是	4	9.76
				否	37	90.24
		G2	优秀率最高比例 (%)	15	1	2.44
				20	1	2.44
				30	5	12.20
				40	1	2.44
				无	33	80.49
		G3	优秀率最低控制分数或档次	85	2	4.88
				90	2	4.88
				A-	4	9.76
				无	33	80.49
		G4	不及格率最高比例 (%)	10	1	2.44
				无	40	97.56
		G5	是否允许突破成绩分布限制	允许	7	17.07
				不允许	1	2.44
				无规定	33	80.49

研究者在编码过程中发现，共有 10 所高校在政策文本中进一步对学生的成绩分布做出了两类强制性或半强制性规定：要求总体成绩符合正态分布与限制优秀率（或不及格率）的比例。有 4 所“双一流”高校明确规定学生成绩应符合正态分布规律；5 所高校要求获得优秀的学生比例最高不超过 30%，各有 1 所高校要求优秀率不得超过 15%、20% 和 40%。为了让成绩分布更加“正态”，部分学校还提出了更为严格的要求，如清华大学规定了获得 A+ 的比例不得超过 5%，北京师范大学则规定了获得 65 分以下者不得少于 3%。

与对优秀学生的诸多规定相比，高校对不及格率的关注程度明显较低。仅有北京大学一所高校规定不及格（60 分以下）的比例不超过 10%，不满足该规定的课程需要进行“调分”以符合学校的分布要求。

与此同时，“双一流”大学对成绩分布限制又保留了一定的调整空间，仅有 1 所高校明确表示不允许突破规定的分布比例限制，其余高校则大多给予了任课教师一定的自主权，但是需要在突破成绩分布限制时提出申请及做出说明。例如，北京大学规定“优秀率超过 40% 或不及格率超过 15% 的课程成绩，须经院（系、所、中心）教学主管领导审核、教务部备案后方可登录”；清华大学的表述是“如有特殊情况，任课教师需向教学主管部门提交情

况说明”。

四、‘双一流’大学学生成绩评价制度反思

(一) GPA 变化敏感度与学业评价適切性之思

通过对“双一流”建设高校成绩评价的政策文本进行分析发现，大多数高校都提供了原始分数与 GPA 之间较为细粒度的转换方案。但是当百分制原始分数被转换为绩点时，转换方法的敏感度有所不同。研究者选取五所具有典型转换算法的“双一流”高校进行比较，从图 1 可以看到，有些高校的绩点转换算法呈现出一条连续性的平滑曲线，学生成绩的每一分变化都精确体现在 GPA 上，如北京大学；有的学校则呈现出区间式的折线形态，处于同一区间的学生共同享有相等的 GPA，位于不同区间的学生绩点至少相差 0.3 分，如中国人民大学、复旦大学。

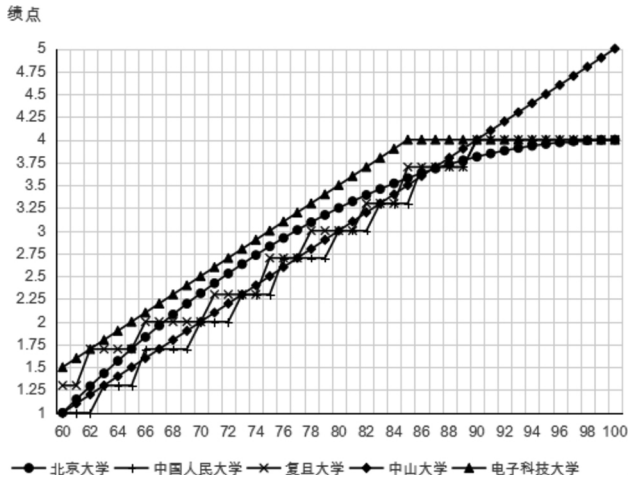


图 1 典型“双一流”高校百分制成绩与绩点转换算法比较

GPA 的首字母 G 所代表的 Grade 一词本身可做等级之意^[19]，意指通过分级对事物进行评价。既然是一种等级制度，笔者认为与原始分逐一对应的高敏锐度的转换体系似乎并不恰当，因为这样与直接采用原始分作为评价标准并无本质差别，失去了转换的意义。另一方面，学生的学业表现无法通过每一分的绝对差距进行衡量，等级制即是一种把简单的连续性分数转化为学业表现的层次性评价方法，让学生明白自身的学业表现处于哪个档次。然而，中国的在学本科生数量已超 1750 万人^[20]，位居全球第一。从评价的角度而言，作为评价对象的在校 生基数庞大，因而也不能简单地采用过于宽泛的分段方法，造成在需要以 GPA 进行学业水平评价时无

法保证合理的区分度。从当前的“双一流”大学实践上来看,绝大多数高校采用的是每5分进一档的GPA转换方式,相对而言,这是一种比较合理的方式,既体现了分级评价的作用,又弱化了原始百分制的精确性评价;但也有部分高校存在过度敏感和较为迟钝的问题,甚至失去了GPA评价制度与原始百分制成绩的差异化价值,有待在后续政策调整中进行优化。

(二) 不同GPA制度与学业表现可比性之思

从之前的分析结果可以看到,当前“双一流”大学所采用的GPA转换方式较为多元,转换算法、“满绩”设定、参评课程范围、是否允许重修已通过课程等都是GPA得分的影响因素。相同分数在不同GPA标准下的绩点存在较大差异,甚至相同成绩能否获得“优秀”的标签也各有不同,产生了高校间“同分不同等”或“同分不同绩”的现象,进而影响到学生GPA表现的跨校比较。清华大学学生之所以对学校的GPA转换方法存在质疑之声,原因就在于该校课程绩点比同类高校平均要低约0.25分,导致学生失去与其他高校的竞争优势。也正是考虑到同样的原因,在国内高校保研与海外高校深造等需要对校际之间学生的学业表现进行横向评价时,并不具有可比性的GPA正在失去它的作用,相对位置与排名成为了考察候选人时更为关心的重点。

当不同高校的学生从本校脱颖而出,进入到新的目标高校寻求深造时,如何真正发挥GPA本应具有的评价作用,在一个可比的制度框架下衡量和评估GPA所反映的学业表现,仍旧是一个颇具挑战的议题。考虑到标准化的考试可以成为高校教学优秀评估和校级比较的指标^[21],且在高等教育大众化的进程中,国家始终是主要的发动者、组织者和监护者^[22],笔者建议在高校拥有成绩评价自主权的基础上,应当由教育部等政府主管部门或高等教育行业组织出台一份具有指导性作用的成绩评价建议文本,基于我国高等教育的实际情况来统筹不同高校之间的评价制度设置,让来自全国各个高校的学生可以拥有相对的可比性,而不致让GPA制度沦为实质上难以发挥作用的一纸空文。

(三) “正态分布”与教师评价自由性之思

从表面上看,由单一的百分制到引入等级制,加以精细的GPA转换方法,我国大学成绩评价制度在不断趋于完善、合理。然而,功利主义心态却无法通过政策文本予以调控,政策文本似乎反而成为

一种可乘之机,令学生整体成绩呈现出上扬的“负偏态”^[23]。这也正是由西方而起的被称作Grade Inflation或GPA Inflation的“分数膨胀”备受关注的直接原因。在中国高校普遍推行学生评教的制度情境下,教师愿意给学生高分以换取学生的满意度。^[24]学生为了获得高分在选课时会将任课教师打分情况考虑在内,使选课在某种程度上成为基于风险规避的理性选择过程,分数则成为这种理性选择的收益。^[25]然而,在成绩普遍很高的课程中,认真勤奋、能力较强的学生很难脱颖而出^[26],高校准确评估学生素质和知识水平的能力受到损害^[27]。

许多高校对“正态分布”和优秀率的限制正是为了避免上述问题发生,强化教师给予学生成绩认定的制度公平性。由于成绩分布相关规定的存在,教师必须更加客观地去评定学生之间的学习成效优劣,也激发了学生在同辈群体中取得更好学业表现的动力。笔者认为这种做法不仅应当被写入政策文本,也应当在实际执行过程中得到遵守,例如可以将教务系统设置为对成绩分布情况进行自动校验。事实上,教师如果采取差别奖励的方式进行学生成绩分配,有助于促进学生学业表现。^[28]不过,当学生的学业表现在整体上超出预期或低于预期时,教师也应当享有对成绩分布进行调整的可能性和自主权。这种自主权的行使应当有明确的政策支持,并且在具体操作时不至于给教师造成行政流程上的过当负担。

(四) 关注高分考生与整体评价公平性之思

就学生自身而言,综合能力、努力程度、同辈素质,甚至人格特质^[29]等都是学业表现的影响因素,这些因素的个体分化造成了学生之间的学习效果和学业绩效差异,也是成绩存在分布的本质原因。同时,由于教育部各类专项计划对中西部、农村、贫困地区的招生倾斜,一些基础薄弱的学生在起步阶段存在较大的学习困难。^[30]高等教育并非只是培养出类拔萃优等生的绝对筛选机制,也应当让每一位学生在高等教育中得以发展,因而学业成绩评定政策需要考虑对全体学生的合理评价,同时兼顾高分与低分学生的学业表现。

在“双一流”大学的政策文本中可以明显发现,与对优秀学生的关注相比,GPA制度对学习困难学生的关注度较低,甚至抹杀了他们进步的可能性:对低于60分或等级为F的成绩,经过转换后的绩点常常统一记载为0分。尽管补考、重修等方式

让成绩不合格的考生不必担心留级和末位淘汰^[31],但这种 GPA 的绝对差距已然造成他们在评奖、评优、保研、出国等竞争中的劣势,并且难以通过后续努力扭转。已经有研究表明促进高分成绩、筛选优异学生或防止“分数膨胀”的做法(如限制成绩分布),会影响成绩较差学生的学业表现以及 GPA 成果^[32];低学业能力的学生很容易受到来自高能力学生的负面影响^[33]。对于这种现象也有高校提出了一些新的主张,例如中国科学技术大学允许学生放弃最多两门课程成绩参与 GPA 评定,南京大学允许学生重修 1 次已通过的课程以获得更高的成绩纳入 GPA 计算。无疑这种做法提高了学生的期望水平,更有利于激发学生的上进心与积极性,有助于提升学生的 GPA^[34],也为在个别课程中表现不优秀甚至存在学习困难的学生留下了“容错”的空间和机会。在未来的成绩评价制度改革中,对优秀学生的遴选与对后进学生的激励应当同时被纳入政策文本制订的考量范围。

五、结语:构建合理的大学成绩评价制度

随着“双一流”高校建设步伐的不断加快,我国高等教育更加开放和自由,GPA 制度也成为我国高校融入全球高等教育体系的必经之路。当高校不断地改进学业评价体系,希望客观、科学地对我国的学生学业成就进行评价时,“分数膨胀”现象反而成为高等教育新的痛点。本文采用内容分析法对“双一流”建设高校学业评价的政策文本进行研究后发现,“分数膨胀”不仅是个体行为的结果,更是制度逻辑之下的产物,GPA 的生产过程有着清晰的制度线索,过往简单的个人利己主义研究视角掩盖了背后的制度逻辑困境。

尽管在具体制度设计和操作层面上“双一流”高校的评价体系不尽相同,但它们都应当发挥出平衡筛选与培育功能,以及促进不同层次学生均衡发展的作用。高分成绩不是高等教育的“指挥棒”^[35],作为学业表现衡量尺度的成绩评价制度不仅应当扮演遴选和排序的角色,更应当发挥出反馈和激励的作用。仍有一些问题需要在将来的制度设计与政策制定中予以更为充分的讨论与合理的安排,例如如何对优秀学生进行合理区分,避免高分的激烈竞争所带来的“分数膨胀”;又如在作为选拔性学生评价制度的同时,如何引入发展性学生评价^[36],在选

课退课、课程成绩放弃、重修等退出和容错机制上为困难学生的激励创造空间等。

诚然,本研究仍有进一步发展的空间。首先,受政策文本可获得性的限制,本研究仅选取了 41 所“双一流”建设高校作为样本院校,对普通院校、高职高专院校和民办院校的制度文本并未进行更广泛的收集与分析。这些院校在生源结构、师资力量以及管理制度等方面均存在差异,今后的研究可以扩大样本范围以探究广泛意义上的中国高等教育学生评价体系。其次,高校教师和学生作为该制度下最为重要的利益相关者,他们对政策文本的评价和对实施效果的内在感知将为研究者带来更深层次的思考,并为回溯到制度层面开展的研究提供认知层面的补充,在后续研究中可以尝试引入质性研究方法以加深对这一议题的洞察力。

参考文献:

- [1] 国家教育部.普通高等学校学生管理规定 (2005) [EB/OL].(2005-03-25) [2019-10-11].http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/200503/t20050325_81846.html.
- [2] Imose R, Barber L K. Using Undergraduate Grade Point Average as A Selection Tool: A Synthesis of the Literature [J]. The Psychologist - Manager Journal, 2015, 18(1): 1 - 11.
- [3] Johnson V E. Grade Inflation: A Crisis in College Education [M]. New York: Springer-Verlag, 2003.
- [4] Bliwise R J. Where are the C's of yesteryear: Indexing Achievement [J]. Duke University Alumni Magazine, 1997, 83(3): 8-13.
- [5] [26] Felton J, Koper P T. Nominal GPA and Real GPA: A Simple Adjustment that Compensates for Grade Inflation [J]. Assessment & Evaluation in Higher Education, 2005, 30(6): 561-569.
- [6] Rosovsky H, Matthew H. Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation [R]. Massachusetts: American Academy of Arts and Sciences, 2002.
- [7] 天津大学档案馆, 天津大学教务处. 天津大学档案馆大数据之 GPA 的秘密 [EB/OL]. (2018-05-31) [2019-05-19]. <https://mp.weixin.qq.com/s/PD0mqAbJd7R8r2XTNwWBYA>.
- [8] 清华大学注册中心, 清华大学教务处, 清华

大学研究生院.“建立促进学生全面发展的学业评价体系”改革方案[EB/OL].(2015-05-15)[2019-04-23].<http://www.doc88.com/p-4962506925441.html>.

[9] Hyde T. Should Schools Implement Tougher Grading Standards? The Economics of Grade Inflation [EB/OL].(2015-09-09)[2019-08-30].<https://www.aeaweb.org/research/should-schools-implement-tougher-grading-standards>.

[10] 张艺璇.深度|谁动了我的GPA?[EB/OL].(2019-04-15)[2019-04-20].<https://twd2.me/wp-content/uploads/2019/04/mygpa.pdf>.

[11] 李丽丽,李江霞.美国高校分数膨胀现象述析[J].黑龙江高教研究,2007(4):38-40.

[12] Gilpin G A, Saunders J, Stoddard C. Why has for-profit Colleges' Share of Higher Education Expanded so Rapidly? Estimating the Responsiveness to Labor Market Changes [J]. Economics of Education Review, 2015(45):53-63.

[13] 邓勇.美国高校学生分数膨胀现象研究[D].长沙:湖南大学,2015.

[14] 刘强.论高校学业考试“分数膨胀”及其应对策略:兼谈高校“等级制”取代“百分制”改革[J].高等教育研究学报,2016,39(2):86-91.

[15][24] 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].社会科学研究,2019,34(1):84-105.

[16] Neuendorf K A. The Content Analysis Guidebook [M]. 2nd ed. Los Angeles: SAGE Publications, 2017: 44.

[17] 邱政.美国研究型大学本科生学业标准研究[D].上海:华东师范大学,2010.

[18] Beckman K. The Top 15 Universities with the Highest Average GPAs [EB/OL].(2018-05-23)[2019-08-03].<https://ripplematch.com/journal/article/the-top-15-universities-with-the-highest-average-gpas-4f4b544d/>.

[19] Cambridge University Press. Grade 1 Definition in the Cambridge English Dictionary [EB/OL]. [2019-10-28].<https://dictionary.cambridge.org/us/dictionary/english/grade>.

[20] 国家教育部.2019年全国教育事业发展统计公报[EB/OL].(2020-05-20)[2020-07-31].[http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202005/](http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202005/t20200520_456751.html)

[t20200520_456751.html](http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/20200520_456751.html).

[21] 田建荣.学业成绩考试:高校教学过程的重要环节[J].厦门大学学报:哲学社会科学版,2002(2):15-18.

[22] 王洪才.大众高等教育论:高等教育大众化的文化-个性向度研究[M].广州:广东教育出版社,2004:265.

[23] 潘懋元.关于学业成绩负偏态分布问题的初步探讨[G]//潘懋元.潘懋元论高等教育.福州:福建教育出版社,2000:469-480.

[25] 彭筱.本科生课程成绩评定中的师生博弈[D].上海:华东师范大学,2019.

[27] Tucker J, Courts B. Grade Inflation in the College Classroom [J]. Foresight, 2010, 12(1):45-53.

[28] Johnson V E. An Alternative to Traditional GPA for Evaluating Student Performance [J]. Statistical Science, 1997, 12(4):251-269.

[29] Farsides T, Woodfield R. Individual Differences and Undergraduate Academic Success: the Roles of Personality, Intelligence, and Application [J]. Personality and Individual Differences, 2003, 34(7):1225-1243.

[30] 宋洁绚.民族院校学生学业成绩评价改革的困境与出路[J].民族高等教育研究,2019(4):11-16.

[31] 陈玉祥.高校学生淘汰制度的社会学分析[J].江苏高教,2009(2):114-116.

[32] Wetzler E L. How Using A Restricted Grading Range Distorts GPAs and Disproportionately Penalizes Low-performing Students [J]. Frontiers in Education, 2019(4):23.

[33] Contreras S, Badua F, Adrian M. Peer Effects on Undergraduate Business Student Performance [J]. International Review of Economics Education, 2012, 11(1):57-66.

[34] Seirup H, Rose S. Exploring the Effects of Hope on GPA and Retention among College Undergraduate Students on Academic Probation [J]. Education Research International, 2011:1-7.

[35] 韩玉志.创新人才培养呼唤高校学业成绩考试改革[J].中国高教研究,1999(6):73.

[36] 刘青.高校发展性学生评价制度研究[D].兰州:兰州大学,2012.

(责任编辑 赖佳)