

海外视野

栏目主持:中国人民大学全民阅读教育研究院院长郭英剑教授

哈佛“限A令”:美国高校分数膨胀的终结信号?

郭英剑

根据日前哈佛大学官方校报《哈佛深红》报道,该校文理学院以458票赞成、201票反对的压倒性多数,通过了一项足以载入其教学史的重要改革:自2027年秋季起,本科课程中“A”等级成绩将被严格限制在班级人数的20%,最多允许增加4个A名额,即所谓“20+4模式”。换句话说,在一个100人的课堂中,获得A成绩的学生原则上不得超过24人。该提案获得约70%的教师支持,被视为哈佛几十年来遏制分数膨胀最激进也最具象征意义的一次制度干预。

消息一出,舆论震动。人们不禁要问:为什么哈佛率先对“A”开刀?

其实,答案很简单。并不单单是因为哈佛“太严格”,还因为一个长期存在且日益严重的问题——分数膨胀,已经迫使哈佛不得不采取行动。

根据哈佛本科教育办公室2025年发布的报告,在2024—2025学年,该校超过60%的本科生成绩为A,而20年前,这一比例仅约24%。换句话说,在短短20年间,A成绩的比例增长了一倍以上,优秀几乎成了“常态”。该报告直言,现有评分制度已无法发挥其关键功能,并在“损害哈佛的学术文化”。

随之而来的问题是,当A变成了一种普遍现象,成绩还能否真正区分学术表现?当成绩失去辨识度,大学又如何向社会、雇主和研究生院证明“卓越”依然真实存在?

分数膨胀的历史

如果说哈佛今天的“限A令”是一次剧烈的制度纠偏,那么它所试图纠正的,其实是一场持续了半个多世纪的慢性病——美国大学的分数膨胀。

今天的人们已经很难想象,在20世纪60年代,美国大学里的A成绩曾经是种十分稀缺的荣誉。

根据长期关注美国高校评分问题的学者研究,在20世纪60年代初,美国高校A等级成绩比例仅约15%,平均绩点(GPA)为2.5。换言之,一个学生获得A意味着处于学术表现的前列。而成绩的功能也主要是“区分能力”,帮助大学、研究生院与雇主识别真正卓越。

事情很快发生了变化。许多研究者认为,美国高校第一次大规模分数膨胀始于越南战争时期。

当时,美国实行征兵制度,而大学生只要维持一定学业表现,即可获得延期征兵资格。这一时期,美国大学平均GPA由2.5升至接近2.9,A等级比例迅速升高,部分研究估计约为30%。有人将这一阶段称为“越战型膨胀”。它并不完全来自学术标准下降,而是在特殊历史情境下形成的一种制度性宽容。教师面对学生可能被送往战场的现实,倾向于给予更高分。这种“同情性评分”后来成为美国大学文化中一种潜在传统。

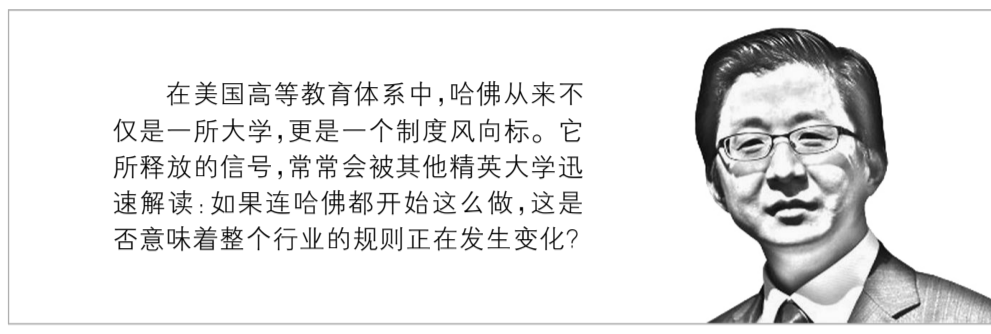
而真正决定美国高校评分体系长期走向的并不是战争,而是市场。

随着美国高等教育大众化,尤其私立高校学费不断上涨,学生越来越像“顾客”,大学则越来越像竞争激烈的服务机构。学校之间竞争排名、学费收入、校友捐赠与就业率,教师则越来越依赖学生课程评价作为晋升与续聘的重要指标。

于是,一个微妙而现实的激励机制出现了。严格评分意味着学生不满、课程评价下降;宽松评分则意味着更好的课堂反馈与更小的教学阻力。

不少美国学者直言,在某种意义上,美国大学逐渐形成了一种“高分均衡”。因为所有人都在放宽标准,所以一位教师很难独善其身,否则便可能被学生评价为“不近人情”或“故意刁难”。

结果是,成绩越来越高,但其意义却越来越小。



在美国高等教育体系中,哈佛从来不仅是一所大学,更是一个制度风向标。它所释放的信号,常常会被其他精英大学迅速解读:如果连哈佛都开始这么做,这是否意味着整个行业的规则正在发生变化?

根据一些学者的长期追踪研究,到2013年前后,美国高校平均GPA已达到3.15,而A等级成绩比例为45%左右。更值得注意的是,私立大学比公立大学更甚,GPA平均高出0.2以上,而精英私立院校则成为分数膨胀最显著的区域。

特别需要指出的是,进入21世纪以来,这种趋势进一步加剧。根据相关研究对1990—2020年美国高校数据的分析,中位GPA增长约21.5%,平均GPA从2.81升至3.15。尤其在新冠疫情期间,美国高校普遍采取更具宽容性的评价机制,包括通过/不通过的评分方式、延期考试、柔性评分以及心理健康优先政策,使分数膨胀进一步加剧。

当人人优秀时,优秀便不再稀缺;当A成为默认配置,成绩也逐渐失去了区分优劣的功能。更深层地看,分数膨胀损害的不仅是成绩本身,更是大学作为“学术共同体”的信任机制。因为评分制度原本是一种严肃的学术判断,它向学生、社会和未来雇主传递了一个信号:什么是努力、什么是能力、什么是卓越。然而,一旦评分越来越像一种情绪安慰、客户服务与制度妥协,大学也就悄然从“标准的维护者”变成“满意度的提供者”。

这也许正是哈佛今天不得不出手的真正原因。

当“A”成为平均分,大学还如何识别卓越

过去,人们通常认为成绩分布应当大致呈现金字塔状:少数顶尖学生获得A,多数学生处于B区间,C则代表基本达标。然而,在今天不少精英院校,尤其私立名校中,这种结构已经发生了根本性变化——A和A-成为“新常态”。部分高校A类成绩占比达到70%至80%,整体成绩曲线明显上移,最终形成一种奇特现象——“顶端压缩”。

所谓“顶端压缩”,指的是成绩大量聚集于高分区间,以至于优秀与良好之间的边界越来越模糊。换句话说,成绩作为评价工具,其“信号功能”正在弱化。

事实上,分数膨胀的后果比人们想象的更严重。

首先,受损的是学生自己。理论上,高分应该是一种激励,鼓励学生更加努力。但当高分变得容易获得,它反而可能削弱学习动机。越来越多美国学者担忧,一种“理性规避风险”的学习文化正在形成:学生不再问“什么课最值得学”,而开始问“什么课最容易拿A”。

于是,“shopping for GPA”(购买GPA)逐渐成为美国大学校园里的流行表达。部分学生倾向于避开评分严格的教师、数学与自然科学等挑战性课程,而转向评价宽松、作业较少、给分友好的课程。

这意味着分数膨胀最终可能悄悄改变学习行为:它奖励的未必是挑战精神,而可能是策略性选择;鼓励的未必是探索未知,而可能是风险规避。这与美国本科教育长期强调自由

探索与学术冒险精神几乎背道而驰。其次,成绩单越来越难以发挥社会识别功能。长期以来,GPA不仅是学校内部评价工具,也是一种社会信号——研究生院、企业招聘者、奖学金评审者据此判断一个人的学术能力与学习品质。而当越来越多的学生拥有接近满绩点的成绩单,这种信号开始“失真”。

不少美国雇主公开表示,仅凭成绩单越来越难以识别真实能力。一些大型企业因此更加依赖实习经历、推荐信、面试表现以及能力测试;研究生院则越来越强调科研经历、写作样本与综合评价。

最后,分数膨胀还制造出新的不公平。表面上高分让所有学生受益,但实际上它常常以不均衡方式分布。在美国高校内部,不同学科、不同院系之间长期存在着显著的评分差异。研究发现,人文社科课程往往比理工类课程给分更高,小班研讨课又普遍高于大型基础课。这意味着,两名同样努力的学生仅仅因为专业不同、选课不同,就可能获得截然不同的GPA。

一个更深远的问题已经在大学之外。一些研究发现,高中阶段已经出现的严重分数膨胀,与后续劳动力市场表现之间存在负相关。因为高分环境削弱了真实的学习强度与知识积累,学生进入大学后,以及大学毕业开始就业后可能更难适应真实的竞争环境。

正因为如此,越来越多的美国教育研究者开始担忧,分数膨胀不仅影响大学内部秩序,更可能制造一种社会层面的“能力错觉”。

哈佛会改变美国大学吗

在美国高等教育体系中,哈佛从来不仅是一所大学,更是一个制度风向标。无论是本科通识教育改革、招生政策,还是近年来的恢复标准化考试要求,哈佛的选择往往具有极强的示范效应。它所释放的信号,常常会被其他精英大学迅速解读:如果连哈佛都开始这么做,这是否意味着整个行业的规则正在发生变化?

因此,当哈佛决定对“A”开刀时,它所引发的问题便不再只是“哈佛要不要限制优秀率”,而是美国大学是否正在重新思考“卓越”的意义。

这一点,从哈佛本科教育专业院长阿曼达·科莱姆的公开表态中可见端倪。她明确表示,希望这一改革不仅改善哈佛自身的学术文化,也能“鼓励其他高校以同样的严谨与勇气面对类似问题”。

换句话说,哈佛并不满足于“独善其身”,而是主动为美国高等教育抛出一个问题:当评分制度已经失灵,大学是否有勇气恢复传统?

从积极层面看,这一改革确实可能产生“涟漪效应”。

首先,它可能推动美国大学重新建立学术评价的可信度。过去几十年,美国高校事实上陷入了一种“高分均衡”困境,因为所有学校都在膨胀,所以没有学校敢轻易收紧。否则,一旦本校学生GPA下降,在研究生申请、奖学金竞争和就业市场中便可能处于劣势。

而哈佛的特殊性正在于,它是少数有能力打破这种均衡的学校。因为哈佛拥有极高的声

誉资本,即便GPA下降,社会也不会因此怀疑其学生能力;相反,严格评分甚至可能被视为质量信号。所以,如果哈佛成功,其他精英大学便更有可能跟进,从而逐渐形成一种新的行业规范,使高分重新具有稀缺性。

其次,这可能改变美国大学评价学生的方式。长期以来,美国高校广泛将学生课程评价作为教师聘任、晋升与绩效的重要指标。这种制度无形中强化了一种现实激励:教师越容易给高分,学生往往越满意。于是,一部分教师逐渐陷入某种“教学消费主义”困境——与其要求严格,不如降低难度;与其制造挑战,不如维持好评。如果评分制度重新变得严格,教师或许更有动力重新聚焦真正的教学问题,即如何提高课堂质量、激发思辨能力,而不是如何避免学生投诉。

话虽如此,哈佛的此次改革能否成功仍有待观察。

首先,根据学生媒体的调查,大约85%的哈佛本科生反对这一提案。他们担心,硬性限制A比例会加剧竞争,削弱合作氛围,并使其更不敢挑战高难度课程。同时,这一政策也会削弱教师的教学自主权。一些学生甚至认为,这会把哈佛重新推回一种“人为制造稀缺”的竞争文化。

与此同时,教师群体却表现出明显不同的态度。在最终投票中,约69.6%的教师支持设立A等级上限。他们认为,当前评分体系已经失真,如果大学无法重新建立评价标准,那么“优秀”终将失去意义,而成绩也将无法承担其应有的教育意义。

其次,之前曾有同类改革失败的前车之鉴。2004年,普林斯顿大学曾推出著名的“分数通胀”政策,试图将A类成绩控制在35%左右。改革初衷与今天的哈佛极为相似——恢复成绩区分度,遏制膨胀。然而多年之后,争议不断出现。一些学生抱怨,在申请与就业市场中,自己相较其他高校学生“明显处于劣势”;而外部机构也并不总是充分理解“更低GPA意味着更严格标准”的道理。最终,普林斯顿大学在2014年终止了该政策。

普林斯顿大学的案例提醒人们,如果只有一所大学独自收紧标准,它可能会因“诚实评分”而遭遇竞争惩罚。这正是在高等教育中的经典“集体行动困境”:所有学校都知道问题所在,却都希望别人先改变。

即便如此,哈佛的改革仍意味着开启一种更深层的发展趋势。因为它真正触及的不是A等级比例,而是美国大学是否仍然过度依赖分数本身。

大学究竟该奖励什么

如果学习成绩只是安慰、激励与满意度管理,那么高分当然越多越好。但是,如果大学仍然相信学习意味着挑战,成长意味着区分,卓越意味着稀缺,那么评分制度便不能彻底失去边界。

虽然这一改革最终能否成功还有待时间检验,因为它既取决于制度设计是否足够细致,也取决于其他高校是否愿意共同回应这一问题,但可以肯定的是,哈佛已经点燃了一场无法轻易回避的讨论的导火索:当“人人优秀”成为常态,大学该如何证明卓越依然存在?

寻回大学教师的耐心

刘爱生

20世纪六七十年代,美国斯坦福大学心理学家沃尔特·米歇尔设计了一个简单的实验:让一群4到6岁的孩子独自坐在桌前,面对一颗棉花糖。他们可以选择不吃,也可以等待15分钟,然后得到两颗。多年后的追踪研究发现,那些能够等待更长时间的孩子,在成年后取得了更大的成就。这个实验揭示了一个朴素的真理:延迟满足的能力,即耐心,是影响人生长期发展的重要心理特质。

大学教师的职业生涯,某种程度上就是一场持续数十年的棉花糖实验。选择“即时满足”的路径,可能是频繁追逐热点,快速产出;选择“延迟满足”的路径,则是深耕一个冷门领域,忍受多年没有产出的“沉默期”,只为有朝一日产出真正的原创性成果。问题在于,今天的大学教师还有多少能像等待两颗棉花糖的孩子一样,守住这份耐心?

耐心是学者成功的心理基石

从心理学角度看,耐心并非简单的“能等”。它是一种复杂的心理能力——当一个人在追求目标过程中遭遇延迟时,会产生不耐烦的情绪;而耐心,就是个体通过转移注意力、重新看待处境、调节情绪反应等方式,对这种情绪进行有效管理的过程。

心理学研究发现,耐心与个体的目标达成能力显著正相关。那些在日常生活中更能保持耐心的人,往往拥有更高的幸福感、更低的焦虑水平和更强的抗压能力。在日常小事上能等待的人,在面对人生重大挑战时更有坚持下去的能力。

在学术研究中,耐心的意义尤为突出。真正的原创性研究往往需要长期投入,例如收集数据、分析数据、反复修改稿件和面对审稿人的质疑等,都需要耐心。更关键的是,学术研究的回报是高度延迟的:一篇论文从投稿到发表可能需要两年,一项研究从立项到产出可能需要5年,一个理论从提出到被认可可能需要数十年。

棉花糖实验的启示正在于此:延迟满足的能力之所以能够影响长期成功,是因为这种能力反映了个体的情绪调节水平、目标专注度以及对长期回报的信任。这些心理特质,恰恰是优秀学者不可或缺的品质。

学术环境如何消解学者的耐心

当前学术环境中的某些文化正在无意间消解学者的耐心。这种消解并非学者个人品质的退化,而是环境因素对个体心理资源的持续消耗。

首先是量化考核带来的“即时满足”诱惑。非升即走、年度考核、聘期评估,这种制度安排将学者的职业生涯切割成一个又一个短期考核节点。每个节点都伴随着明确的回报:发表几篇论文,获得多少奖励。在这种环境中,等待意味着风险,如果3年内没有足够产出,就可能面临淘汰。当制度本身将短期产出确立为重要目标时,学者选择“一年磨一剑”而非“十年磨一剑”,恰恰是一种高度理性的适应行为。

其次是失败的高成本削弱了人们等待的意愿。在当前的学术评价体系中,论文被拒、项目申请失败、研究未能产出预期成果,这些“失败”的代价是累积性的——它们直接转化为下一轮考核的劣势。当等待的终点可能是“两颗棉花糖”而是“什么也没有”时,谁还愿意等待?

最后是社交比较的加剧加剧了心理资源的消耗。学术社交媒体的普及、论文发表速度的透明化、同行之间“比着发文章”的氛围,使学者们时刻处于被比较的状态。看到同龄人每年发5篇论文,自己却还在为一个数据挣扎,这种持续的心理压力会消耗宝贵的情绪调节资源,使耐心变得难以维系。

如何帮助学者重建耐心

寻回大学教师的耐心,需要从个体调适和环境支持两个层面入手。

从个体层面看,学者需要重建对“延迟满足”的心理信任。棉花糖实验告诉我们,孩子愿意等待的前提是“相信两颗棉花糖真的会出现”。同样,学者愿意等待的前提是“相信长期投入终将获得回报”。面对短期产出的压力,学者可以有意识地提醒自己:这项研究的真正价值不在于它能发表在哪儿,而在于它能解决什么问题;这个领域的深耕需要时间,但时间终会证明它的意义。

从环境层面看,学术机构需要为耐心创造心理安全空间。如果制度只奖励快速产出,学者的耐心就会被系统性惩罚。当学者感知到“等待”本身就是制度缺陷所导致的不利处境时,任何个体的心理调节都将事倍功半。因此,高校需要建立更包容的评价机制。在聘期考核中区分不同类型的学者,设立长期研究项目,允许学者在数年内无产出但有进展,在评审中重视研究的深度而非数量。这些制度设计能够降低学者对“延迟”的焦虑感,为其耐心的维持提供心理空间。

当然,这里需要澄清的一点是,强调耐心的重要性并不意味着所有研究都应该慢下来。面对突发公共突发事件、亟待突破的技术瓶颈时,研究的“速度”本身就是一种价值。此外,学术共同体需要重塑对“失败”的认知。当一个学术社群能够接纳“研究进展不顺利本身就是研究的常态”时,学者对“无产出时期”的焦虑感就会下降。

棉花糖实验的后续研究还发现一个有趣的细节:那些能够延迟满足的孩子,并不是天生“能等”,而是他们更善于让等待变得不那么难熬——他们会唱歌、玩游戏,会把注意力从棉花糖上移开。这个细节揭示了一个根本的真理:耐心不是天生的禀赋,而是可以被习得、被训练、被支持的心理能力。

大学教师的耐心同样如此。它不是对制度压力的消极忍受,而是在明知环境不利的情况下,依然有能力保持内心的平静、专注和对长期价值的信任。寻回这份耐心,既需要学者个体主动地进行心理调适,也需要学术共同体为这份调适提供可持续的空间。毕竟,真正的学术创新从来不是“一年磨一剑”的产物,而是那些愿意等待“两颗棉花糖”的人的成果。

(作者系浙江大学政治与公共管理学院教授)

中国大学评论 AI时代预言文科是“春天”抑或“冬天”重要吗 尤小立

AI时代预言文科是“春天”抑或“冬天”重要吗

尤小立

作为“类劳动力”的人工智能(AI)出现以来,大学似乎又找到了努力的目标。全面迎合AI浪潮看上去是踩中了时代的风口,但它与“完全拒绝”的保守态度一样,都是在走极端。受此影响的“文科冬天论”和“文科春天论”也不例外。

“文科冬天论”的预言者多非文科出身。他们只是松散的组合,不过是在网络上的升学规划、考研、就业指导的间隙以及非正式场合发表些有关文科命运的悲观言论。在他们看来,AI时代,文科将面临淘汰。他们主要是根据AI迅猛发展后劳动力的需求走向作出判断,符合当下社会的整体趋势以及面对此趋势的诸多想象和推论。学生普遍偏向选择广义的理工科,以及国内外大学缩减文科专业的现象也仿佛成了一个现实佐证。如果说,以前的社会曾有“学好数理化,走遍天下都不怕”的流行语,如今如今的流行语或许就是“得AI者得天下”了。

当然,随着AI本身的不断进化,类似的极端观点也在自我修正。因为AI要替代的是整个人类,而所谓人类,就不包括文科出身者,理工科出身者也不能幸免。所以,当国际著名AI大厂传出招收大量文科背景员工的

信息,就立即被一些文科出身者挂上网络,四处传播,为它们所支持的“文科春天论”正名。

实际上,这样的需求纯粹是被动性的。它是一种“被需要”,就像一个零件被嵌入整个理工科及其应用体系,其本身并没有成为主体的可能,只是为机器、为产品服务,充其量是间接为人类社会服务。

与“文科冬天论”不同,“文科春天论”大多出自大学校园中的文科专家。他们乐观地预测说,在AI时代,文科不是面临淘汰,而是迎来了一个巨大的发展机遇。

然而,这类“春天论”并不是建立在雪莱所谓“冬天来了,春天还会远吗”的经验推理上,而是停留在“前现代”的想象中。比如,在他们看来,文学艺术往往是闲散的产物,当AI全面接管社会生产、劳动之后,人们的闲暇工夫多了起来,文学艺术也必然成为追求的目标。殊不知当生产文学艺术产品的任务也被AI所接管,AI工具能够通过自我想象与自由组合,自动设计并生产文学艺术产品之后,人只能成为被动的欣赏者、点评者、消费者,或者被AI审美所塑造的对象。此时的人们很可能会像前些年围棋世界冠军输给美国AI围棋程序AlphaGo之后逐渐疏离围棋运

动一样,失去对文学艺术的创作兴趣。

表面看,“文科冬天论”是面向现实的,但实际上,这一见解充满了学科偏见和视域盲区。看不到学科之间相互依存的关联性,以及文科诸学科的不可替代性,是源于学科之间的知识隔膜,也是原有知识积累不足、知识面过于狭窄,步入职场后又缺乏跨学科阅读的热情所致。而“文科春天论”似乎是建立在理想基础上的,但理想主义者绝不是脱离现实的,否则理想就会沦为幻想。停留在“前现代”的想象中只是“文科春天论”的一个具体表现,长期空泛议论也让他们养成僵化的思维定势,进而影响了对现实事物的判断力以及捕捉现实问题的敏感度。原本AI冲击下的文科命运是一个极其现实的问题,但到了他们那里就成了“上天入地”的根本性话题,仿佛非要到“古今中外”绕上一大圈,才能显示其深刻与博学。

从理论上说,现代社会如果要正常运转,需要通过现实主义与理想主义的博弈达到一种动态平衡。但是,“冬天论”和“春天论”不仅没有肩负起这一社会责任,反而衍生出了伪问题。

因为“二论”的预言者并没有深入了解AI,这样就形成了一个奇特的现象:不懂得

AI,不了解算法的人在预言AI的影响,而懂得AI、了解算法的人却在一门心思进行技术和产业竞争。由于彼此没有真正意义上的思想交集,前者的言说也不可能对后者产生任何启示。

只要对“大学”的基本概念稍有了解,就不难看出,大学要不要文科本身就是一个假问题。在大学发展的长河中,文科也像其他学科一样经历过高峰与低谷。这样的波动在大学历史上可谓司空见惯,AI出现前就存在,AI时代再度出现也不至于让人手足无措。

进一步看,即便没有AI,由于整个社会的技术化倾向,以及由此产生的急功近利,大学文科早已处在边缘化的境地。以上两种调性相反的预言,就是在这个背景下产生的,它们的共同点是对大学文科失去信心。然而,由于类似的预言脱离AI场景,彼此之间无法争出高下,也不可能提供明确的答案,只能在论争中获得一时的存在感。

因此,与其进行无谓的预言,不如将此问题转换成如何让文科更好地适应AI时代。这显然更有价值,也更有利于为大学管理提供有理念、有章法,并符合大学实际的“AI赋能”方法。

(作者系苏州大学政治与公共管理学院教授)