

# 发展改革动态

2021年第9期共138期

发展规划处

2021年5月25日

## 【聚焦评价改革】

### 董云川：高等教育普及化语境下的评价反思

摘要:就中国高等教育的质量建设而言,评价有功,但绝非万能,当高等教育进入普及化时代之后,教育评价的方向、方式和方法亦需相应调整,必须规避盲区,明确标的,分辨因果,调整重心,评判有据,谨慎作为。既知其可为,亦知其不可为。面对有生命的教育活动和动态的院校行为,必须知道有的评不了,有的不能评,有的无需评,有的小心评。去繁就简之后,教育评价方能够规避副作用而更好地践履其推动教育进步的使命。

关键词:高等教育普及化;教育评价;评价盲区;评价重心

#### 论点摘编

□ 近年来,伴随着越来越重要、越来越强劲的教育评估角色与功能的演进,针对高等教育系统中错综复杂的院校发展、学术研究以及人才培养问题,“该不该评”“能不能评”“评什么”以及“怎么评”等一系列专业化评估的基本问题并未得到善解。教育评价看似简单,其实纠结,虽早有共识,但高校内外的疑虑始终有增无减,迄今尚待理论界予以更加深入明晰的回应。

□ 进入普及化时代,高等教育到了专注充盈教育组织与教育个体生命力的时候了。科学的教育评价目的在于不断激发教育组织的生命力,时刻唤醒师生成长的皈依感,持续笃定学者潜心探究的真情怀!舍此无他。而这一点,恰恰是当下教育评价的短板。

□ 因此,评价发生的动机到底源于外因还是内因,评价到底该由他主还是自主,学校到底是被动接受还是主动选择,评价促进了压力还是动力,是水到渠成的质量检验还是缘木求鱼的行为规制,迎评促建的姿态是伪装的还是本能推动的?这一系列问题都很重要,不能被忽略。

□ 教育评价与学术评判不能够简单等同,否则,教育评价过程中就少不了生出形式霸权,造成评判模糊,再加之仲裁机制的缺位,纠偏的可能性极小,评估的副作用就会大幅度衍生。

□ 换言之，“一流”有别，事实与打算不一样；“建设”亦有别，在建与建成不一样；“进入一流”也有别，判断与臆断不一样。仅从评价用语时态上看，已经、即将、计划、打算、念想，统统不能同日而语。

即便在教育资源稀缺的时代，当受教育者在选择要不要接受教育的时候，隐性的、个别的、非制度化的评价就同步而生了，也就是说，乙（受教育方）对甲（教育方）所提供的服务自有判断。对不对、该不该、能不能、值不值的观察和抉择与教育自身的发展步伐如影随形。

而在教育资源逐步积累、教育机会不断增加、教育竞争迅速加剧的情形下，甲乙双方对彼此的评判要求也相应提高，伴之以社会发展及分工细化，所谓的第三方应运而生，随即，“丙”的评价与制衡作用自然凸显并在现代教育系统中扮演着越来越重要的角色。

现代意义上的高等教育起源于欧洲中世纪，推进于工业革命，兴盛于 20 世纪。发达国家在 20 世纪下半叶相继进入了高等教育普及化时代。之后，由学者马丁·特罗建立的一种解释性框架获得了广泛认同，在中国备受青睐，以至于超乎作者本人预料，成为了政府教育改革与发展决策的主要参照系。于是，在诸多教育政策文本的制定或在讨论教育改革与发展实践问题之时，“精英”“大众”“普及”等相关字眼成为了超高频率使用的关键词。

与相关的话语方式同步，中国高等教育的发展步伐也迅即在最近的四十余年，自豪地、以跨越式的姿态从“精英化”迅速迈向“大众化”并如愿进入了“普及化”的大时代。相应地，数量与质量的矛盾同步衍生，高等教育发展的质量隐患逐步显现，质量保障的呼声也越来越高，无论是政府还是民间，对于高等教育品质的诉求和质疑成为热点和焦点。

近年来，伴随着越来越重要、越来越强劲的教育评估角色与功能的演进，针对高等教育系统中错综复杂的院校发展、学术研究以及人才培养问题，“该不该评”“能不能评”“评什么”以及“怎么评”等一系列专业化评估的基本问题并未得到善解。教育评价看似简单，其实纠结，虽早有共识，但高校内外的疑虑始终有增无减，迄今尚待理论界予以更加深入明晰的回应。

本文不再赘述教育评价之“熟知的常识性表达”“不容分说的政策话语”“诠释性的学理论证”以及“教科书式的功能陈述”，只想结合中国教育改革语境，基于教育质量现实，紧扣高等教育普及化的发展趋势，挖掘一下教育评价的根性逻辑，并基于教育的主体性角色，深入反思一二，以供同道商榷。

### **首先，评价有盲区，要谨慎地厘清教育评价的标的**

当教育资源的供给从紧缺到富足，越来越外显的评价活动推动高等教育陷入盲区。

其一，教育的生命感不停地被削弱。大学组织本来是一个活态的生命体，具有自组织调节机能，其置身于特定的社会环境之中，建立、发展、壮大，自有其生存之道。好大学，好学科，好教育，好学者，好学生，好成果离不开科学的评价，但根本上并不是评出来的。

当下盛行的话语方式似乎颠覆了这个认知。鲜活的教育生命被网格化的指标分割殆尽，动态的教育系统被静态的指标框定，大学中的芸芸众生无不对标对策，寻找活路。远的不说，就看看千校一面的大学网站首页，听听众口一词的校长演讲内容，端倪毕现。

其二，教育情怀逐渐淡漠，教育情义迅速流逝。师生在日常教育或学术活动中本应深切体验到的神圣感和意义感现在普遍淡化。以空洞无物或无情无义的指标为鹄的的评价活动交叉往复，

把人们的注意力迅速地从教育的内涵引向了外延，评价活动可以有效甄别大学及其师生人人身上悬挂的标签，独独评不出教育情怀的价码，判不了教育情义的斤两。因此才会有学校调侃说：“我们离世界一流大学只有十米远了。”

进入普及化时代，高等教育到了专注充盈教育组织与教育个体生命力的时候了。科学的教育评价目的在于不断激发教育组织的生命力，时刻唤醒师生成长的皈依感，持续笃定学者潜心探究的真情怀！舍此无他。而这一点，恰恰是当下教育评价的短板。

当高等教育从少数人特享的专利摇身一变成为普罗大众可以平等分享的权利之后，教育评价促进教育发展的大方向和质量标的是否越来越清晰且合理？愚以为，高等教育评价在宏观层面的理想标的如下：

宏观层面的理想标的：

要让教育更像教育，具有更纯粹的文化品格；

要让大学更像大学，成为更加专业化的学术组织；

要让老师更像老师，愈加彰显学者风范；

要让学生更像学生，孕育更多探奇求新的莘莘学子。

如此衡量，中国高等教育的现实状况及其教育评价的效益显然不尽人意。否则，就不会有那么多的过来人和学者不停地回望“精英化”时代的高等教育生活了。

中观维度的健康标的包括：

教育评价的结果能否适时反馈到质量的改进环节，有效促进教育多样化的发展，扩宽精英时代的单一口径，造就立体的教育生态，促使教育从一元转向多元，助力不同的院校去实现各得其所的目标。

实际效果看来远未达到预期，当下高等教育的发展诉求愈加统一规范，中国大学组织越来越缺乏个性。

微观角度的適切标的无疑要与“人才培养”这个高等教育的根本任务紧密关联，而人才的品质维系务必落实于师生并体现于师生。大学理应成为师生六大修炼的场所，以实现个体“身体、心智、智慧，策略、手段、机会”的全面发展。前三是本，难以评估，后三为末，易于甄别；前三稳固，才能持续发力不懈怠；专注后三，难免昙花一现而短寿。高等教育不能舍本逐末。

**其次，评价有立场，要理智地辨明教育评价的理由**

所有的评判行为都是有立场的。对教育供给方行为的评价，起始于受教育者，发展于教育投资者（政府或民间财团），兴盛于第三方专业机构的确立。

仅从表面上看，作为被评价的学校组织及教育行为，在这样的关系系统中似乎处于被动地位，被要求，被约束，被指点，被筛选，最终所争取到的，无非被接纳的命运。

然而，从本质上看，教育评价行为的生发无不依托并服务于教育目的本身。如果没有旺盛的教育需求以及教育的规模，教育评价行为就不会产生，而其一旦不能够合理地服务于教育的健康发展，则教育评价行为就将失去存在的意义。

因此，评价发生的动机到底源于外因还是内因，评价到底该由他主还是自主，学校到底是被动接受还是主动选择，评价促进了压力还是动力，是水到渠成的质量检验还是缘木求鱼的行为规制，迎评促建的姿态是伪装的还是本能推动的？这一系列问题都很重要，不能被忽略。

进一步分析，教育评价的立场和姿态决定了教育评价的话语方式。在强大的政策表述中，忽而要求大学“应该如此”“必须遵守”；忽而要求大学“不得这样”更“不得那样”；一下子鼓励，一下子严禁；今天要“规范”明天要“破除”；前十年“惟此”后十年“惟彼”。

作为教育运行的主体，高等学校何以适从，破与立的科学逻辑何在，谁说了算？如此这般折腾之后，原本具有周期性、长期性运行特征的高等教育体系，以及具有自组织调节功能的高等学校被来自于教育外部忽左忽右的政策体系所掣肘，畏首畏尾，不知所措。结果导致校校、人人把原本该投入到学术探究和人才培养中的宝贵精力耗费在了应付外部名目繁多的评价检查活动中，于是就演绎出了世界一流的、轰轰烈烈的迎评促建运动。显然本末倒置了。

从学术维度分析，支撑教育评价话语的理由亦需要考证质疑。什么“对”什么“不对”？当前一些官样学术文章把“教育评价”的功效描述得神乎其神，无以伦比，似乎好学校、好学科、好学者统统都是通过评价体系甄别和遴选的结果。

其实，评价既有边界更有局限。教育评价与学术评判不能够简单等同，否则，教育评价过程中就少不了生出形式霸权，造成评判模糊，再加之仲裁机制的缺位，纠偏的可能性极小，评估的副作用就会大幅度衍生。

教育评估专家常常通过浮光掠影的材料审阅辅之以二到三天的校园巡察，就可以对一所大学的教育学术和人才培养说三道四，指点江山，振振有词地指出被评院校的错误与缺陷，圈点出弱势与短板，断言特色与创新，肯定自己认可的，否定与自己不同的。其合理性依据及可行性逻辑相当值得怀疑！

大学是一个超复杂的文化系统，历史基础不同，人文地理不同，经济状况不同，学科结构不同，师资队伍不同，学生来源不同，学校要生存要发展，自有考量，外人隔靴搔痒，何以轻易做出论断而干扰到学校的命运。

再说，教育活动和学术状态虽有量化内容，但更多的部分只能够模糊评论并不适合做精准评价。无论是综合评估还是专业评价，均涉及估价与研判的可行性和科学性问题的，这绝不是小问题。估价者的话语表述往往天马行空，诸如，方法不妥，逻辑不严，研究不深，形式不完备，论据不充分，结构欠妥，观点不明确等，似是而非，反正说多说少都“不错”，但其实大家都不明确“如何才对”。

而研判者则有必要对论点正误进行辨析，指出观点对错的理由和结论科学与否的缘由，还要对所谓“不妥”的教育行为进行矫正商榷，同时通过内涵评判讲明“为什么”必须如此。何其容易，更不用说动辄调度成百上千的专家们窜来窜去，交叉评审，其中不乏颐指气使的非内行“砖家”和虽在学校讨生活但其实根本不懂教育的“南郭先生”。

### **第三，评价有时态，要端正教育评价的因果关系**

教育质量不是评出来的，也不是从天上掉下来的，更不是凭主观想象而得的。教育生长的状态终究离不开学校组织及其师生个体的点滴言行和持续作为。对于聚沙成塔的教育所做出的评量到底是凭借主观推测还是依靠客观判断，这个问题并不简单，亦非空穴来风。

现如今，许多人张口闭口言必称普及化高等教育时代即将是一个“高质量发展”的时代。这个论断不知从何而得？其用语时态尚值得认真推敲。任何一个时代都是良莠并存，泥沙俱下的。比如，大凡提到“博士”一词，一般泛指三种，一种为已读博士，一种为在读博士，一种为想读博士，彼此大不相同。

面向未来，社会进步加速，人工智能挑战，疫情阻隔变幻，中国高等教育是不是能够自然进入，或者按照主观时间表如期进入一个高品质、高质量、大发展的时代，也同样不是依据口号、意愿和猜测就可以达成的。显然更不可能简单地通过现实的评估就排定出未来一流大学、一流学科、一流学者于某一时刻在世界上的位次。能够进入、计划进入与梦想进入大相径庭。

换言之，“一流”有别，事实与打算不一样；“建设”亦有别，在建与建成不一样；“进入一流”也有别，判断与臆断不一样。仅从评价用语时态上看，已经、即将、计划、打算、念想，统统不能同日而语。

与此相应，当人人都在畅谈一流的时候，我们到底在谈什么？万物因果相连，如果人们只谈未来的目标和预期的一流结果，而闭口不谈或者极少谈实现一流目标的原因和条件（亦即高等教育的办学规律）的话，与痴人说梦也相去不远。

教育部长明确指出了“四个回归”的正确发展方向，教育评价组织及其专业化的活动本该以此为准绳，大力促进教育行为的有效回归，而非相反。教育回归常识无疑是正确的归因方式，那些在漫长的教育历史长河中得到公认的真理教育命题始终应该成为教育评价活动的指南。

比如，“没有爱，就没有教育；只有爱，也没有教育”，这是对教育动机的拷问；又如，如果“一分耕耘，一分收获”的话，跨越式的教育发展跳过了哪些环节，舍弃的环节经得起质疑吗？再如，“最好的教，就是让学生学会学；最好的学，就是让学生给别人讲”，岂不早已契合了新时代教与学的翻转趋势。还有，“一把钥匙开一把锁”，传统教育训导如此，未来因材施教亦属必须，而当下评估所好之变来变去的教育模式难道不值得怀疑吗？再者，“教育科研是做出来的，而不是写出来的”，类似教育研究中理论与实践的常识性逻辑是否在评价中被颠倒了，所以大家才不得不细数成堆的立项课题来衡量一所学校教育的优劣。最后，“名师名校无法速成”，无疑是对现行匆忙慌乱的教育发出的预警，中国高等教育独缺的，正是对教育生长与学术孵化周期的耐心等待。

上述列举的种种常识，是教育活动因果律的典型写照，恰恰是这些基本的常识性命题构成了教育“结果”的“成因”。类似常识是否在教育行动中得以彰显实施，并在评价过程中得到了首肯和保护，是衡量院校质量评价活动科学与否的关键。

也就是说，当结果既定为事实，评价“结果”的意义其实并不如想象的那么重要，而通过合理的评价，沿着教育的因果律，努力将注意力吸引至“原因”的挖掘和改变上来，才是教育评价发挥积极作用的最优化选择。

换言之，教育评价要努力去做“对的”事情，而促进学校去把事情“做对”。借助专业评价让学校去改变“原因”，而无需耗费师生大量的精力去迎接一批又一批的“评估专家”，应对填写教育学术活动之外一堆又一堆的表格材料。因果关系颠倒连带的形式主义评估活动，劳而无功，弊大于利。

#### 第四，评价有阶段，要适时调整教育评价的重心

万物生长有节律，社会现象也不例外，从无到有，从少到多，从简到繁，从轻到重，在不同的发展阶段，每一件事情的轻重与优劣都是相对的。要做到科学合理就必须通过适时而适度的重心调节，才能够有效契合不同发展阶段的不同需求，教育发展因之而健康有序。

以道家三生万物的阶段来划分，人类的社会生活与生产方式无非在“三”的逻辑里循环往复，大有规律可循。人类生活起初“简朴”，然后越来越“繁琐”，到后来以追求“简约”为品质；社会时尚先从“新奇”发端，然后引发“从众”潮流，继而以显摆“奢侈”为标的；在生活消费方面，初期拮据而“入不敷出”，随后占有且“挥金如土”，后来知道要“环保生态”方为高尚；世相人生免不了从“小众的个体”出发，之后努力挤进“合众的群体”以求认同，最终发现只有“大众的个体”才能够如鱼得水；而文化选择的步履，起初必须以他者的“马首是瞻”，随即彰显个人以“自鸣得意”，最后拥抱自由从而“相互欣赏”；以上诸例终而复始，归于大道自然：先自然，继而摆脱自然，最后回归自然。

对应辨析高等教育的生长，无外乎早期的精英化，其后的大众化，今天的普及化。每个发展阶段的重心不一样，因此，每个发展阶段的质量评价关注点也应该不一样，进入高等教育普及化阶段之后，评价的重心有必要依据相应逻辑渐次调整。

重心位移的目的，当然是为了更好地服务于高等教育的健康发展。精英阶段的教育评价重在资源，大众化阶段的教育评价重在公平，而到了普及化阶段，教育评价的重心当该重在内涵。所谓高等教育的内涵，知易行难，现代高等教育历经千年发展史，尚在不断探索，各奔其途，中国高等教育历经百年磨难，仍在孜孜以求，未得善解。

进入新时代，高等教育的内涵品相，政策导向是“科教兴国，办学方向，立德树人”，学理表述是“大学理想，科学追求，教育情怀”，个体实践则离不开“生命体验，好奇探求，成长故事”。

在普及化教育时代，谋大谋小也成了问题，贪多求大的取向可以适度放下，小众个体的培育应该得到更多关注；动静取舍也需要权衡，静态的业绩指标可以铺陈宣讲，但鲜活的生命故事需要深入发掘；繁琐还是简约亦需要调整，大道至简至易，逻辑愈简，覆盖面越宽，评价方法越简单副作用越小；评价的内外力量协同作用尤需关联，外部评价应该偏重于“问责”，而内部评价方着重于“品质”；运动式评价需要大幅度减少，规避形神两分的评价，实质意义上的学术反省自查以及学术共同体间的互动交流应该大力加强；术与道的重心更需不断调适，重机巧的评估技术要适度，智慧型的质性研讨应该得到凸显。概言之，不同的阶段说不同的话，轻重缓急不能同日而语，教育评价理当伺机转换对策。

综上所述，根据事物演进的逻辑，教育评价活动在高等教育普及化阶段当然还会大有作为，但真的很有必要调整战略战术，从而使得教育评价的作为更加合理、科学，更加有意义、有价值。

未来，纯粹定量的办学状态数据完全无需人工介入，只要做得到及时、真实、客观公正地发布，社会大众可以据此做出各自适切的选择，则善莫大焉；纯粹静态的教学科研指标也无需专家参与劳心费神，由中介机构和工作人员罗列比对即可，只要能够尊重事实，客观公示，则不同院校的教育运行状态也就一目了然；唯其在动态而错综复杂的人才培养生境和学术研究领域，才需要在学术共同体之间谨慎开展专业化的甄别商榷，促进平等交流，实现友好互动，方有可能得出相对合理的评价结论或建言对策，而接受与否，亦必须由受体学校自主裁定。只有本学校、本学科、本学者，才是自己教育质量和学术生命的终决者。

积极的评价活动，引导教育活动由表及里；消极的评价活动，促使教育由内而外。教育普及了，迫切需要还给高等学校更多自主生长的空间，让师生更多地专心于教育本身才是评价的本分和圭臬。彼时，如果非要他者来评估，最明智的选择还是少评。（原文刊载于《江苏高教》2021年第4期第1-5页）

作者简介：

董云川：云南大学高等教育研究院研究员、博士生导师。