



借鉴与参考

第 2021 (01) 期

# 研究生教育发展动态

## 【主题】

高等教育评价体系创新

## 【导语】

评价之于当今中国高等教育改革与发展的意义，已不仅仅在于其作为质量保障制度的核心构成要件和推动质量提升的政策工具，而是一个关联到整个事业健康发展的改革攻坚难点，具有极为复杂的系统关联性，它的破与立、成与败都具有牵一发而动全身的意义。

## 【观点】

- 01 学者正当行使学术权力对高等教育评价改革至关重要
- 02 高等教育增值评价：一种绿色教育评价方式
- 03 教育评估视域下的高等教育发展
- 04 大学评价的“中间道路”
- 05 破“唯论文”应以贡献作为学术评价的取向

# 导语

“

2020年10月，中共中央、国务院印发了《深化新时代教育评价改革总体方案》，其中对高等学校评价、学科评估、国际交流合作评价、高校教师科研评价以及大学生学业评价等，都提出了改革的方向以及相关原则性要求。针对评价议题以如此高的规格出台文件，可谓罕见。这足以表明：尽管经过多年的探索实践，目前我国各类教育评价制度与体系得以逐步成型，但其存在明显的短视与功利取向，已经对高校发展、人才培养、理论创新以及国家需求的重大技术突破带来了某些不良效应，因此教育评价的改革势在必行。与此同时，它也意味着：评价之于当今中国高等教育改革与发展的意义，已不仅仅在于其作为质量保障制度的核心构成要件和推动质量提升的政策工具，而是一个关联到整个事业健康发展的改革攻坚难点，具有极为复杂的系统关联性，它的破与立、成与败都具有牵一发而动全身的意义。

对于高等教育领域各种评价存在的问题，其实多年以来学术界并不乏关注与讨论。但是，在业绩至上的治理框架以及计量思维和指标主导的评价氛围中，各种理性乃至反思性与批判性的声音都显得极为羸弱。如今“总体方案”的出台，无疑为人们全面检视其弊病、寻求破解之道提供了难得的机遇。不过，客观而言，任何制度一旦形成路径依赖甚至形塑为人们的一种思维惯性和习性，破不易立更难。因为在破与立之间，横亘的并非简单的工具和技术是否合用的问题，如指标设置合理与否和测量方式妥当与否，而是涉及观念以及理论层面众多深层次且需要正本清源的问题。譬如如何理解大学作为学术组织的属性，如何理解学术创新活动的本质特征及其所需要的环境和条件，究竟如何认识学术国际影响力的内涵，如何破解评价中人的能力结构潜隐性与呈现方式多样性困境，如何看待高校与政府、社会三者间的关系，如何理解不同社会力量介入及其激励机制与高校创新活动间的关系，如此等等。

上述问题都带有本源性，反映了学术活动的本体价值与大学组织运行的内在逻辑，也是我们正确理解与审视高等教育评价活动或不妨称之为元评价的基本理据与价值原点。也唯有把这些问题弄清楚，在各方之间特别在管理者中形成理论共识与共同的价值体认，我们才会理性审慎地处理好诸如是否有必要评、评什么、谁来评、如何评等第二位的问题，才有可能走出各种评价项目繁复冗杂，评价过程工具化、指标化和技术化的误区，穿越破与立之间的各种迷障。

——阎光才 华东师范大学高等教育研究所教授

”

注：此期内容摘自或节选《苏州大学学报（教育科学版）》2021年第2期和“苏州大学学报”微信公众号

# 学者正当行使学术权力 对高等教育评价改革至关重要

文 | 陈廷柱 华中科技大学教育科学研究院教授

教育评价改革是全面深化教育领域综合改革的关键所在，高等教育评价改革是教育评价体系改革的重中之重。贯彻落实《深化新时代教育评价改革总体方案》等文件精神，肯定会加大主观评价的频度、广度与力度，学者们能否正当行使学术权力直接关系到高等教育评价体系改革的成败。

## 一、高等教育评价改革与学者正当行使学术权力密切相关

近年来，与高等教育评价改革相关的政策文件不断强调要建立“代表性成果”评价机制，加强同行评价。2016年教育部颁布的《关于深化高校教师考核评价制度改革的指导意见》指出：完善同行专家评价机制，积极探索建立以“代表性成果”和实际贡献为主要内容的评价方式，将具有创新性和显示度的学术成果作为评价教师科研工作的重要依据。2017年中共中央办公厅、国务院办公厅印发的《关于深化职称制度改革的意见》指出：建立以同行专家评审为基础的业内评价机制，注重引入市场评价和社会评价。随后，教育部、中共中央办公厅、国务院办公厅、科技部等多个部门出台的相关文件，以及《深化新时代教育评价改革总体方案》都反复提到完善同行专家评价机制，重视专家评议意见，特别是要突出同行专家在基础与前沿研究领域的作用。这说明国家决心走出过度依赖外显性或量化指标评价的偏颇，通过扭转评价导向和优化学术生态，以落实立德树人根本任务和扎根中国大地办学。

教育评价若不以分数或升学率、论文数与期刊级别、项目经费与立项单位等客观指标为主要依据，或者说要建立代表性成果评价机制以突出质量、创新或服务导向，必然要依赖同行专家才能完成评价任务。高等教育领域的评价任务尤其与高深知识的传播、生产与应用密切相关，推进高等教育评价改革更需要发挥同行专家的作用，因为只有学术同行对相关活动、成果或学者的真实水准与实际贡献最具评价权。学术同行在评价学术事项方面所表现出来的参与权、判断权与表决权，笼而统之，也就是所谓的学术权力。相对于行政权力，社会各界往往更加认同学术权力。然而，实践中学者群体并非孤立地行使学术权力，受权力部门过度介入、人情社会、学术标准缺失等内外部环境影响，很可能导致评价改革政策所倡导的同行评价出现异化现象。更有学者指出：学术系统内部总是存在一个等级分布的结构，现实中的同行认可权力运行往往并不民主，西方历史悠久的“同行评议制度”事实上不过也是不同权力博弈的一个最不坏的途径。因此，在看到高等教育评价改革在加持学术权力的同时，也要提出和审视学者能否正当行使学术权力的问题。否则，高等教育评价改革可能会陷入内卷化的境地，不仅于事无补，还会徒增一些解释、论证或引证材料填报任务。

## 二、学者正当行使学术权力面临的挑战

**其一，评审任务的学科跨度。**伯顿·克拉克（Burton Clark）曾指出：“即使是最通用的学科，也包括本学科大多数人所共享的观点与专门学说，并以此将本学科与非本学科研究者区分开来。”这就是说，学科是具有边界的，某个学科的学者不大可能了解其他学科的情况，不适合承担其他学科的评审任务。在学者个体层面，其学科知识储备与学术判别能力更为有限，能够有效进行学术评价的事务并不多。2019年中共中央办公厅、国务院办公厅印发了《关于进一步弘扬科学家精神加强作风和学风建设的意见》，明确强调“科研诚信是科技工作者的生命”，并规定科技工作者“不参加自己不熟悉领域的咨询评审活动，不在情况不掌握、内容不了解的意见建议上署名签字”。但在实际工作中，我们要经常面对其他学科的科研项目、人才项目与各种奖项的评审材料。一些高校或政府部门在委托学者进行同行评议时，也经常是将多个学科的评审材料直接打包发给评审专家。若此，看似是同行评价，其实是外行评内行。即便是相近相关学科，学者们也未必胜任评审任务。

**其二，评审专家的时间成本。**高质量地评价送审的学术成果，不是一件轻而易举的事情，需要评审专家认真阅读申报材料，甚至是反复比较各个研究成果。学术事项或成果评审需要投入足够的时间，因此学者正当行使学术权力就面临着时间成本问题。必须承认，确实有很多评审专家能够认真对待评审任务，愿意投入时间和精力，以保证评价结论的科学性与公正性。但是，也不排除因为各种因素的限制或干扰，确有部分评审专家只是简单浏览一下有关申报材料，便会快速给出评审结论。所以，在提交各种由同行专家进行评审的材料之时，大家都有一个基本共识，即如何才能快速抓住评审专家的眼球并打动他。总之，时间成本是影响评审质量的重要因素。即便是评审专家擅长的学科领域，若不能投入足够的评审时间，也是对学术权力的亵渎。

**其三，评审专家的自我保护。**评审专家隶属于不同单位和学科圈子，基于所在单位与学科领地的利益，因此有自我保护的需要或倾向。我们在校内经常参加各种评审，首要的任务就是为本单位、本学科的申报人或候选人争取入围的机会。学校与次级跨院系的各种学术委员会委员，可以说主要是各自所在学院与学科的代言人，他们讨论或评价学术事项也很难超越所在学院与学科的利益。此外，评审专家出于自我保护的考虑，也往往采信某些更为客观的标准，比如论文数量、刊物级别、转载引用情况等。说到底，破“五唯”所针对的那些指标或事项，到了同行评审专家那里，仍然是决定性的判据。原因是，如此操作既简单也容易给出令人信服的解释，忠实地行使学术权力反而成为出力不讨好的事情。若此，很多评审任务即便增加了同行评审环节，也难以达到质量、创新与服务导向的评价改革预期。

**其四，修复评审生态有待时日。**出现“五唯”局面非一日之过，破除“五唯”顽疾也非一日之功。凡事皆有过往，一旦形成某种生态，处理起来就非常困难。可以说，我国同行评审的学术生态并不理想，以至于大家宁愿相信那些客观的标准或指标，却对教育评价改革方案倾向于依靠专家学者主观判断的做法或办法，持着怀疑或谨慎的态度。尽管“五唯”顽疾确实到了不改不可的地步，但是学术生态问题或许比“五唯”顽疾所造成的不良影响更为严重。我们看到，教育部、科技部以及有关基金管理部门也在大力整治学术生态，然而学术生态修复谈何容易。



## 三、学者如何才能正当行使学术权力

以上所谈到的挑战，涉及各个层面与方面，特别是与学者自身及评审生态有关的内容，通过出台一些办法与措施，是很难立马见到成效的，确实需要坚持久久为功。但是，在力所能及的范围内，为规范和约束学者正当行使学术权力，仍然有努力为之的空间。

**其一，为正当行使学术权力提供规则与技术支持。**在规则供给方面，尽快完善学术同行参与评价的制度约束与流程控制，参照国际同行在不同学科中开展评价的成熟经验进行本土化的制度改造；强化源头治理，对各种评审的组织或管理单位实施更为严格的问责制度。在技术支持方面，加大各种评审平台的共建共管共享力度，切实避免申报、评审活动在小圈子中循环往复，让所有的申报材料与评审信息在技术平台上均有痕迹可以追溯倒查，进而通过评审大数据筛选出真正具备专业素质与精神的学者群体，并匹配到相应的评审任务。

**其二，充分发挥专业学会在各种评审中的作用。**专业学会是某个领域学术同行的“无形学院”。有学者指出：正是这种“无形学院”的组织形式决定着科学知识的发展。某个专业领域的学术成果由所在专业学会组织或参与评价，这对于被评的学者个人也相对具有说服力。一些国家在相关政策设计上注重发挥专业学会的评价功能，我国也在向一些从政府机构分离出来的学会组织转移评估评审任务，但众多专业学会与评估评审任务无缘。专业学会是学者共同体，是重要的评价资源，政府、高校与各种评审平台若能与专业学会携手合作，或许可以为高等教育评价改革趟出一条新路。

**其三，积极培育专业精神，拒绝参与不当评价活动。**学者们是最具批判精神的，然而同时也是最为保守的群体。政府在力推教育评价改革与净化学术评审生态方面，可谓是用尽心苦，并频繁出台了各种文件与举措。扪心而问，学者们对“五唯”顽疾的形成难辞其咎，对破除“五唯”顽疾也责无旁贷。不少学者为了个人或单位利益，对非专业领域的成果贸然进行跨界评价，找人打招呼、为人打招呼、听人打招呼的现象屡见不鲜，严重危害了学术评审生态。在此，建议学者们在进行同行评价时，要从学术共同体的声誉出发，秉承专业认知与专业精神，拒绝参加非专业领域的评审活动，严肃对待、公正处理、全力投入自己比较熟悉的专业领域的评价任务。政府主管部门或高校在委托学者群体开展同行评议时，也务必尊重学者的评价结论，并做好评审信息公开服务，建立有效反馈机制，推进阳光评审，使同行评价接受必要的监督。行有不得，反求诸己。坦率地讲，高等教育评价改革的成败，关键在学者群体是否普遍具备专业精神。而培育专业精神的关键也在学者群体自身，愿学者们以高等教育评价改革为己任，正当行使自己手中的学术权力，为繁荣我国的学术事业做出应有的贡献。



# 高等教育增值评价： 一种绿色教育评价方式

文 | 蒋 凯 北京大学教育学院教授

教育我国高等教育评价历经了从不完善到走向完善、从粗略到规范的发展过程，评价思想也逐渐从注重最终结果和绝对成绩的“资源观”向注重发展过程和相对提高的“增值观”转变。高等教育增值评价强调评价的改进功能和激励功能，兼顾导向功能、鉴定功能和调控功能。探索高等教育增值评价，并不是要将其取代以往的高等教育评价方式，而是力图将这种过程性、发展性评价与以往注重结果的评价有机地结合起来，构建更加科学合理的高等教育评价体系。

教育评价改革是全面深化教育领域综合改革的关键所在，高等教育评价改革是教育评价体系改革的重中之重。贯彻落实《深化新时代教育评价改革总体方案》等文件精神，肯定会加大主观评价的频度、广度与力度，学者们能否正当行使学术权力直接关系到高等教育评价体系改革的成败。

教育评价具有导向功能、鉴定功能、改进功能、调控功能和激励功能。在教育评价的上述功能中，鉴定功能是指判断被评价对象合格与否、优劣程度、水平高低；改进功能是指促进被评价对象为实现理想目标而不断改进和完善行动；激励功能则是指激发被评价者的成就动机，激励他们努力学习或工作，取得更大的成就；等等。尽管上述功能在传统教育评价和现代教育评价中都得到了不同程度的体现，但相对而言，**传统教育评价比较注重评价的鉴定功能和调控功能，现代教育评价则强调评价的导向功能、改进功能和激励功能。**

**教育增值评价是一种过程性评价、发展性评价。这种评价蕴含了新的理念。首先，教育增值评价是一种绿色理念的评价。**教育增值评价与绿色GDP的理念有相通之处，它以推动教育均衡、高质量、和谐有序的绿色发展为核心关注点。教育增值评价关注学生学业成就的高质量增长，尤其是学生学习动机、方法和能力等的有效增长，兼顾学生人文素养等柔性指标的提升，体现出和谐有序的绿色发展理念。**其次，教育增值评价是一种可持续发展理念的评价。**学生接受教育是一个动态发展的过程，随着他们融入学校环境的程度不断加深，学业表现和综合素质提升随着时间变化而变化。教育增值评价并不以某一特定时间点、某次特定测试的结果作为判断标准，而是将学生从一个起点到终点的过程性变化加以综合考量，测量学生能够长久维持的提升状态。这种增值评价方式既满足了诊断学生学习状况的需要，也不会由于过于强调甄别和筛选功能而损害学生的学习热情和学习效果，因而体现出可持续发展的特征。

在高等教育领域，与传统的总结性评价相比，**增值评价的特点在于关注学生的多次学业表现，强调过程评估，注重发挥评价的改进功能和激励功能。**与总结性评价相比，**高等教育增值评价还注**

**重从纵向维度提供翔实的数据，以供办学者、高校、院系和教师有针对性地分析和改进教育质量；注重使用新的高级统计方法，以更加准确地测量学生发展增值和教育效能。**

在学生评价方面，高等教育增值评价已经发展出了横向评价法和纵向评价法等方法。国外已经发展出若干比较有影响的高等教育增值评价工具，如美国的全美大学生学习投入调查（NSSE）、大学学习成效评价模式（CLA）、大学学术能力测评（CAAP）、高校毕业生技能评价（GSA），澳大利亚的学术能力和进步测试（MAPP）等。高等教育增值评价的对象不限于学生，还包括对教师、学校、院系和学科等的评价，阿斯汀曾指出对课堂教学、教学绩效等的评价可以采用增值评价方式。

近年来，增值评价逐渐进入我国教育评价之中，这一绿色、可持续发展评价方式在高等教育领域中得到了一定程度的运用。例如，华中科技大学沈红教授团队承担国家自然科学基金项目“高等教育增值研究”本科生能力测试，对全国16个省份83所高校本科生的批判性思维能力、创造力、人际交往能力和问题解决能力进行实证研究，分析本科生批判性思维能力增值情况并对其影响因素进行探讨。国内还有机构将增值评价理念引入高校教师评价中，以改进以往我国高校教师评价侧重鉴定、甄别和筛选功能，忽视改进和激励功能的不足。

**我国高等教育评价历经了从不完善到走向完善、从粗略到规范的发展过程，评价思想也逐渐从注重最终结果和绝对成绩的“资源观”向注重发展过程和相对提高的“增值观”转变。**在20世纪80年代中期至今国家教育行政部门组织的历次高等教育评估中，增值评价理念逐步得到了加强。从最近两轮一级学科评估看，国家教育行政部门逐渐重视高等教育增值评价，加大对教学和人才培养的重视，不再只侧重师资、资源投入和科学研究。不同的高等教育质量评价观，实际上反映了不同的高等教育质量观。高等教育质量评价“资源观”是一种静态评价观，重点评价高校，关注师资队伍、设施设备物质条件，遵循学术标准；高等教育质量评价“增值观”是一种动态评价观，评价对象以高校学生为主，发挥多元主体评价的智慧，关注学生发展特别是学生接受高等教育前后的变化，兼顾对高等教育领域其他方面的评价。



# 教育评估视域下的高等教育发展

文 | 胡钦晓 曲阜师范大学教育学院教授

20世纪后半叶，伴随人力资本理论获得普遍认知，教育尤其是高等教育之于国家发展的重要性日益凸显。教育目的是否达到、教育组织孰优孰劣、教育经费如何分配等，不但缠绕着高等教育场域内部中的组织和个人，也关涉到高等教育场域外部的利益相关者，而这些问题又都是紧紧围绕高等教育评估提出的。因此，教育评估与高等教育发展日渐成为高等教育理论研究者与实践探索者的重大关切话题。

## 一、高等教育评估的重要性

从教育起源来看，教育与评估可谓是相伴而生。在甲骨文中，“教”字的右边是一个人执着教鞭，左下方是一个“子”字，左上方是两个交叉符号，有人认为这表示鞭打的痕迹，我们认为应该是《易经》中的“爻”字，表示交替的意思。甲骨文中的“爻”字，与“教”字左上方的两个交叉符号相同。因此，“教”并非仅表现为惩罚的一面，也具有奖励的意蕴，是惩罚与奖赏交替进行的过程。在中国古代社会中，教育也绝非一味惩罚。《说文》中强调“教，上所施下所效也”，《论语》中强调“循循然善诱人”“不愤不启，不悱不发”等，都说明教育充满着人性关怀。退一步来说，无论是鞭打的痕迹，还是惩罚与奖励的交替，都表现出教育与评价的密切关联。

从中国高等教育来看，各类教育评估引发广泛争议。2020年10月，《美国新闻与世界报道》(U.S.News)公布了2021年世界大学排行榜，将山东省两所高校的数学学科推向了风口浪尖。在国人看来，一所省属院校的学科超过国内顶尖高校，于情于理都是不被认可的。事实上，超过日常认知的评估结果远不止这些，只是因为人们或是选择性无视，或是完全无视造成的。在U.S.News的同年排行榜中，清华的工程学以100分的成绩，排在了麻省理工学院(MIT)之前。再看英国《泰晤士报》(THE)的排名。在THE的2021年世界大学排行榜中，中国的北大和清华均有一个学科进入世界前十，分别为北大的工程技术学科和清华的教育学科。清华大学而非北京师范大学或华东师范大学的教育学科进入世界前十，无疑脱离了中国人的日常认知。北京大学的工程技术学学科能够进入世界前十，同样也经不起仔细推敲。在2019年“软科世界一流学科排名”中，北京大学在工学的22个分支学科中，无一进入“冠军榜”。根据教育部第四轮学科评估结果，排名前10位的工科中也无北京大学。可见，无论是“软科世界一流学科排名”，还是教育部第四轮学科评估，北大的工科在国内都不属于顶尖的，遑论世界范围内的比较。就北大自身而言，其工科尽管发展迅速，但与其强势的文理学科相比，还当有不小差距。我们列举这些实证，无意于否认清华、北大之于中国大学的顶尖地位，更无意于否认清华之教育学、北大之工程学的学科实力，而是强调要理性认知各类教育评价，切不可沦为梁启超所憎恨的“旁观者”，也不可沦为鲁迅所讽刺的“看客”，更不可沦为阿伦特所批判的“群氓”。



## 二、高等教育评估的复杂性

**从教育评估的主体来看**，既有内部评价，也有外部评价；既有政府评价，也有中介评价；既有国内评价，也有国外评价。可以说在当今社会，任何组织和个人都能对教育说三道四、指点一二。

**从教育评估的对象来看**，圣贤如孔子者，尚且为自己对学生的评价感到后悔，更何况评价对象是被伯顿·克拉克（Burton R.Clark）称为“有组织的无政府状态”（organized anarchy）的大学。

**从教育评估的指标来看**，由于评价主体的多样性、评价对象的复杂性、评价数据的模糊性、评价成本的高额性等原因，评估指标的选择就如同组织决策制定，同样甚至更为复杂和艰难，少数简约的评价指标无异于组织决策制定中的“垃圾桶理论”（the garbage can theory）。

**从教育评估的方式来看**，尽管存在少数的主动性评估，但更多是属于全然不知状态下的被动评估；尽管存在组织和个体在场的评估，但更多是属于组织和个体缺席的评估。

**从教育评估结果的运用来看**，有的组织直接声称无视各类商业性评估，有的组织选择性地运用有利于自己的评估；有的国家直接将评估结果与财政拨款挂钩，有的国家打着公平的旗号进行“猪肉桶”（pork barrel）式的利益均沾。所有这些，都为高等教育评估的复杂性，蒙上了更为多样、更为模糊的面纱。

评价指标不合理？  
缺乏独立客观的评价机构？  
评价结果应用不理想？



## 三、高等教育评估的热点问题

当下中国高等教育评估纷繁多样，其中又以“双一流”“破五唯”引发民众广泛关注。前者主要是针对大学和学科的遴选和评估，后者主要是针对教师的晋升和考核。

梳理“双一流”建设高校的遴选和评估，不难发现具有以下基本特点：

**(1) 自上而下性。**2017年1月，教育部等印发了《统筹推进世界一流大学和一流学科建设实施办法（暂行）》，9月21日，就发布了“双一流”建设高校名单。无论是一流大学建设高校还是一流学科建设高校的遴选，都带有较强的政府指令性。

**(2) 身份继承性。**其中一流大学建设高校基本上承袭了原来的“985工程”高校，一流学科建设高校基本上承袭了原来的“985工程”和“211工程”高校。

**(3) 广泛质疑性。**2020年是“双一流”建设的中期评估期，各高校纷纷组织专家论证建设成效，鉴定结果一度引发广泛质疑。先期进行中期评估的高校，获得了从“全面建成”到“整体实现”，从“完成任务”到“达成度高”等不同程度的肯定性评价，可谓皆大欢喜，但是群众并不买账。舆论压力之下，后续的院校或学科中期评估，有的不敢拿出定论，有的则主动承认不足，表现出明显的谨小慎微。

**(4) 圈内循环性。**无论是从“211工程”到“985工程”，还是从“985工程”到“双一流”建设，均表现出明显的圈内循环。这种圈内循环已经严重影响高校、教师及学生的身份认定。如同“割韭菜”般的高考招生录取，伴随名牌高校推荐免试研究生的份额扩大，使圈内循环愈演愈烈。这种圈内循环性，无论是从高校社会资本、学术资本的生成来看，还是从学术自由、教育生态的发展来看，都是极为不利的。



2018年10月，科技部、教育部等五部门联合发出通知，强调“破四唯”，“四唯”即唯论文、唯职称、唯学历、唯奖项；2018年11月，教育部办公厅下文，强调“破五唯”，增加了一唯，即唯帽子；2020年10月，中共中央、国务院发布《深化新时代教育评价改革总体方案》，提出克服唯分数、唯升学、唯文凭、唯论文、唯帽子的顽瘴痼疾。如果将以上结合来看，可以得出“破七唯”。事实上，现实教师评价中，所要破的“唯”还有很多，尤其是“唯项目”。严格说来，科研项目是一种风险基金。且不说从历史上看，许多创新性成果是在不经意间产生的，就从当下中国来看，谁也不能确保国家重大招标课题就一定会比国家重点或一般课题的成果质量高，有的看似重大课题，到最后很可能变成一种闹剧，“汉芯一号”就是最好明证。因此，相比较其他“七唯”，“唯项目”的不可预期性、不可测度性、不平等性会更大。此外，如果不能“立”，而先强调“破”，也会带来更大的无序。教师考核中，缺失了硬性指标约束，权力寻租的空间也会随之增加。

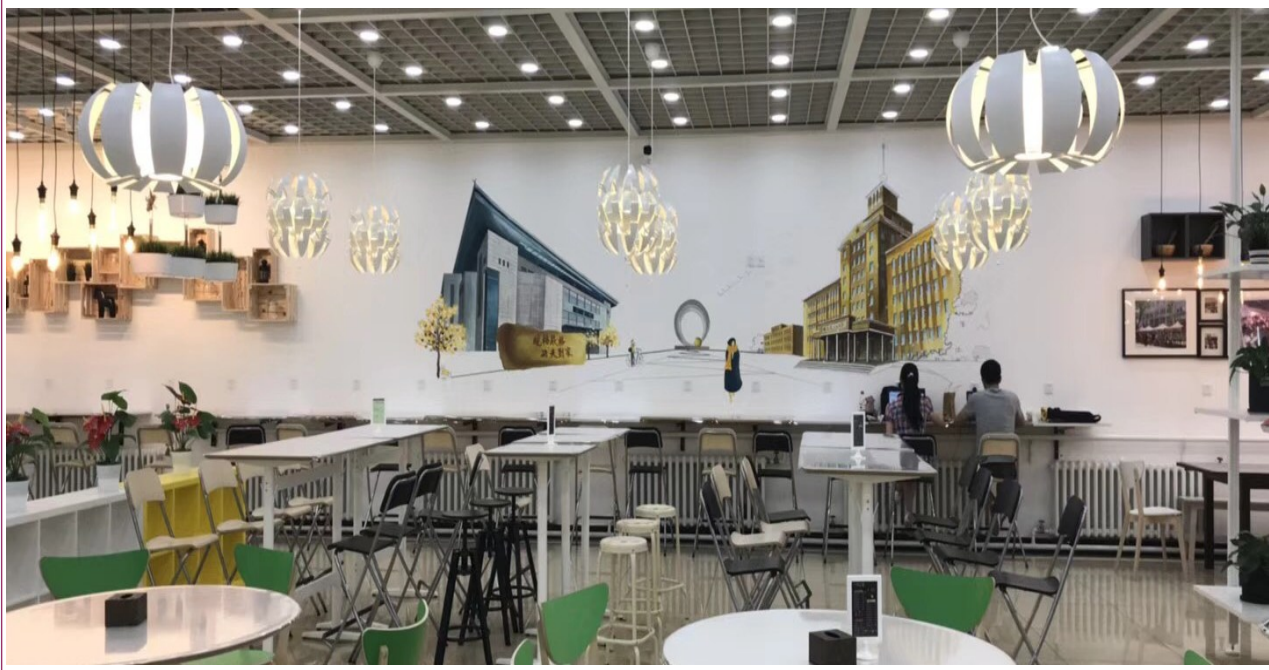
#### 四、高等教育评估的若干思考

**大学以及学科发展要开放竞争。**无论是组织还是个人，在一个公正、公平的开放环境下进行竞争，才能够促进组织或个人的良性发展。井底之蛙或夜郎自大式的封闭循环，对于大学、学科、教师、学生的学术发展皆是有百害而无一利。

**高等教育评估要淡化政府色彩。**中世纪以降，政府关心和支持大学发展，是高等教育强盛的重要保障。但是，无论是学术自由还是教授治学，都是高校学术发展所必需的组织特点。大学只有摆脱政府“襁褓”式的关怀，才能走向自立、自强。这些都需要建立在政府给予高校充分信任的基础上才能够真正实现。

**各类教育评估要适当延长周期。**十年树木，百年树人，教育发展有着自身规律。从世界一流大学的形成来看，大多是自然生成的结果。学者创新需要一定时间的闲暇，学科成长需要持续性的积累，一流大学形成更需要传统的积淀。

**完善成果产出导向的多维评价。**由于教学具有润物无声性，科研具有高度或然性，因此，借鉴专业认证中的“成果产出导向”，或许是未来高校、学科、教师以及学生评价的基本走向。此外，从“四唯”“五唯”走向“多维”，也或许是在打破旧制度之后，如何建立新制度的一个逻辑进路。



# 大学评价的“中间道路”

文 | 王建华 南京师范大学教育科学学院教授

实践中过度强调大学自身的理念性，或将希望寄托于理想的大学评价体系都是不现实的。关于大学的评价唯一可行的或许就是寻找并践行一种“中间道路”。

受新自由主义、计算主义和管理主义的影响，当前大学评价面临三重困扰：**要不要评、评价的方法、评价结果的使用**。真实的大学评价既不像新自由主义者、计算主义者和管理主义者所说的那么美好，也不像反对者所认为的那么不堪；同样，真实的大学作为一类社会组织其生存与发展的境况既不似量化评价和问责制的支持者所指责的那么低效、保守和自利，也不像反对者为之辩护的那么理想和完美。我们既不能以大学的理想来反驳现实中的评价，也不能以评价的理想来苛责现实中的大学。“过度与不及是恶的特点，而适度则是德性的特点。”

## 一、在评与不评之间

我们时代的大学需要评价，甚至必须接受评价有其客观必然性，不是由大学人的主观意愿决定的。具体而言，一方面，在知识经济和创新创业全球化的背景下，大学作为地区、国家和全球创新体系的轴心机构，对于经济和社会发展起着至关重要的作用。另一方面，大学评价作为更大的新自由主义政策议程的一部分，在调节和建构大学与政府、产业之间的关系方面起着不可或缺的连接作用，已经成为国际组织、国家、地方政府、大学联盟以及大学的管理者“治理”大学的重要工具。**某种意义上，当前对于大学的评价是大学的所有利益相关者的理性诉求，而非外力的强加。**对于外部利益相关者而言，通过评价可以监督或督促大学改进自身的教学和研究工作，以更好地服务于国家的战略需求和社会需要。对于内部利益相关者而言，评价的结果不但可以彰显大学之于经济社会发展的重要性，而且可以有力回应外部利益相关者对于大学的问责，进而有助于大学从政府和社会获得更多资源。

我们时代虽然从政治、经济，甚至高等教育自身发展的需要上讲，大学评价有其必要性，但经验表明，大学评价的后果并非总是积极的或正面的。相反，随着评价的增多甚至是滥用，一些意图之外的负面后果正越来越让人对于大学评价的正当性和合理性产生怀疑。评价作为一种大学治理和治理大学的工具，带有“一刀切”式的强制性。无论是政府主导的还是第三方的评价，一旦被纳入政策议程都会对大学的知识治理产生深刻的影响，“特别是大学正在产生的知识类型，正在为谁产生这些知识，以及如何评估知识的传播”。极端情况下甚至会产生“评价的悖论”，即根据政府以及第三方的评价结果，很多大学在评价的各方面都表现优秀，但实质上可能既没有生产出新知识、培养出创新型人才也缺乏服务经济社会发展的真实能力。

鉴于此，大学评价必须审慎考虑其适用范围或边界，避免滥用评价。换言之，“非必要，不评价”。大学作为一类专业组织，绝非所有的活动都必须接受政府或第三方的评价。对于人才培养和科学研究等核心事务，大学自身理应拥有评价的自主权。即便是那些适合于或需要外部评价介入的事务，评价也绝非越多越好；更不意味着只要是外部的或第三方的评价就一定是科学的、公正的、合理的，并要求大学必须接受或认可。如果我们承认在后工业社会中大学之于经济社会发展具有极端的重要性，就必须承认大学相对于政府和社会的独立性。不自主则无责任。一个完全依赖型或依附性的组织（大学）很难成为后工业社会的轴心机构。

## 二、在测与不测之间

除过于频繁和花样繁多之外，当前大学评价中另一个备受诟病之处就是偏好量化。某种意义上，对于统计和量化的偏好不是大学评价的独特问题，而是我们时代所有评价的通病。只不过与其他领域的量化评价相比，其对于大学的伤害更加严重。在量化评价的驱动下，“一所大学早已忘记了大学的功能，且野蛮地发展着，毫无协调理念或现实制约”。与早期的评价强调“专业判断”不同，当前的大学评价更注重技术层面上的“统计”和“测量”。如果说以前对于大学优劣的“判断”需要专业知识，那么当前的大学评价其合法性和科学性则主要基于统计学。根据统计学的原理，“一旦某个现象被量化标准转化，它就可以进行加减乘除，即使这些运算在现实中没有任何意义。数字会带来一种让人舒心的虚幻，无共同尺度的东西也可以拿来衡量比较，因为数学计算总是‘有办法’的……数字会生造一个根本不存在的通用统计基准”。实践中由于理性被简化成了计算，所谓大学评价考察的仅仅是大学的可量化的数据与指标体系的相关性，而非大学发展及其影响因素间的因果性。对计算主义者而言，大学评价中测量和统计必不可少，因为没有统一的指标根本无法客观比较大学的优劣。为了满足评价中对测量的需要，大学不得不放弃或暂时搁置自身特殊的组织目标和价值取向，降格为一个普通的知识和人才“生产组织”。伴随着基于测量和统计的评价实践的蔓延，大学像商品一样被公众“评头论足”。在简单的数字和等级面前，大学以及关于大学的评价实现了“祛魅”，大学的发展过程被作为符号的“排名”或“等级”所支配和控制。更高的排名更强的竞争力既引领大学追求卓越也刺激着政府和社会加大投入。无论政府、社会还是大学都会反复援引量化评估的数据来为其决策进行辩护。最终大学评价既成为一种象征性的权力，亦难以避免地沦为一种“符号暴力”。

大学评价中“指标的陷阱”和“测量的迷思”客观存在。事实上，无论基于何种先进的方法或复杂的技术，我们都不可能实现对于大学的精确测量，并基于测量进行科学的评价。无论如何，也无论何时，大学评价中的统计和测量都只能衡量一些最简单的事实而无法反映大学在现实世界中的真实状态。大学事关人性的启蒙和塑造，涉及文化的传承与知识的创新，绝非那些简单的科研指标、师生人数或产业收入等可以衡量。那些基于测量的评价所揭示的或展示的只是“片面真相”“主观真相”，甚至是“人造真相”，而非“事实性真相”。对于大学评价我们需要更宽广的、综合的视角，我们需要承认高等教育系统的复杂性和多样性，更需要承认大学评价的“不可测性”。这种“不可测性”与其说是大学评价工作面临的困境，不如说是大学作为一个学术性的专业组织的优点。大学评价绝不能只关注那些容易测量的简单事实，而忽略了其他极其重要的事项。公允地说，对于大学而言，绝不存在唯一的精确的关于质量或水平的标准。当然，这也不意味着因为大学评价中有不可测量的事项而完全否认基于测量对大学进行评价的任何可能性。实践中大学评价既要努力去测量那些可以科学测量的，也要考虑使用测量之外的其他方法对大学里那些极其重要但又不可测量的部分做出专业“判断”。换言之，大学评价中应坚持“可测，则测；不可测，则不测；可测与不可测之间的不要强行测量”。

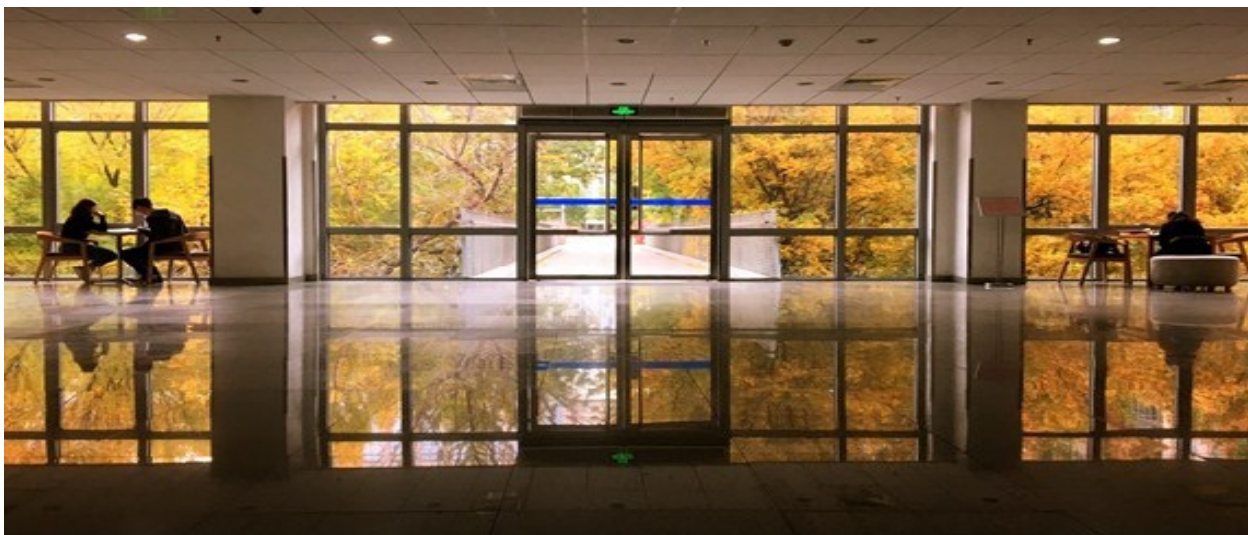


## 三、在用与不用之间

在“评与不评”和“测与不测”之外，关于大学评价的另一个重要问题就是对于评价结果的使用。近年来，受管理主义的影响和学术锦标赛的驱动，高等教育中以测量为基础的绩效评价以及基于绩效评价的问责愈来愈流行。评价的结果不仅事关大学的排名、声誉，而且直接影响资源配置。为了能够在各类评价中占据有利位置，评价指标成了大学及其利益相关者关注的焦点。无论大学还是大学人为承认而竞争的不再是某种稳定的学术地位，而是变动不居的在各种评价中的“表现”。大学需要年复一年地应对各种各样的评价，并力争在所有的评价中有好的表现。结果就是，测量这条“尾巴”已然开始摇动大学这条“狗”。换言之，大学评价日益成为高等教育改革和发展的“指挥棒”。

但问题的关键在于，很多时候基于测量的评价所指引的方向未必符合大学的常识理性和高等教育发展的规律。很多时候所谓的评价既不是对于大学发展真实状态的“事实陈述”，也不是“专业判断”，而只是对于某些数据的“统计”。由于指标的选择或指标体系的不同，大学评价的结果通常具有可逆性或不确定性。此时如果强行将某种量化评价的结果作为对于大学进行问责的依据乃至唯一依据，势必诱使大学为了符合量化评价的指标而偏离正常的发展轨道，从而既不利于大学自身的可持续发展，也不利于大学促进经济社会发展。事实证明，“随着我们建立一个更加严苛的责任架构，我们需要弄清楚我们在计算什么，为何如此计算，谁选择了衡量标准，衡量标准又是如何选择的，衡量者和衡量标准之间有何政治联系，以及这些衡量标准会带来什么样的激励机制”。作为生产高深知识和提供“高等的教育”的制度性场所，大学存在的理由和发展的目标包括但绝不限于人才培养、科学研究和社会服务。从中世纪至今，经过近千年的发展，现代大学已经成为一个高度综合性的机构，大学的存在和发展涉及现代社会的方方面面，评价结果往往“牵一发而动全身”。鉴于大学的极端复杂性和高度重要性，任何一种评价都不可避免地存在这样或那样的局限性，对于评价结果的使用稍有不慎就极易引发连锁反应和未意图的负面后果。

基于此，对于任何一种大学评价结果的使用，尤其是当与资源配置相挂钩时，我们都要慎之又慎。实践证明，好的问责制会诱致大学发展的良性循环；坏的问责制则将会导致大学发展的恶性循环。面对大学评价的结果，我们需要思考：这种评价方式及其结果在高等教育领域是否具有可接受性？是不是不实行这种评价高等教育相关领域的发展就会受到抑制？这种评价在发挥有效作用的同时有没有不可避免的副作用？如果实践证明这种评价方式及结果在高等教育领域不可接受，大学应明确地拒绝；如果这种评价可以接受但对于高等教育发展并无显著促进作用，大学可以选择忽视或无视；如果实践证明该项评价是有效的或至少是利大于弊的，大学则需要积极应对因这项评价而带来的市场竞争和政府问责。





# 破“唯论文” 应以贡献作为学术评价的取向

文 | 张东海 华东师范大学高等教育研究所副教授

2018年，习近平同志在全国教育大会上提出要扭转不科学的教育评价导向，克服唯分数、唯升学、唯文凭、唯论文、唯帽子的顽瘴痼疾。此后，“破五唯”成为各类学术评价改革的共同取向，不管是教育部开展的破五唯专项清理行动，还是其高校教师职称制度改革指导意见，抑或是第五轮学科评估中体现出的基本导向，都是贯彻落实习近平全国教育大会讲话精神的具体措施和纠正过往学术评价不良导向的实施方案。

在高等教育领域“破五唯”的各项具体措施中，“五唯”所指的对象各有不同。据罗燕考察，作为学术评价负面清单的“破唯”始于2016年中共中央颁发的《关于深化人才发展体制机制改革的意见》，其中列出唯学历、唯职称、唯论文三项，习近平在全国教育大会上针对各级教育提出“五唯”，教育部关于清理五唯专项行动通知中的表述又有所不同，“破唯”的队伍继续扩大，仅高等教育领域就包括了学历、资历、帽子、论文、项目、奖项、职称等七项，这些都是高校学术评价中司空见惯的评价依据或评价对象，在评价实践中相互交叉，形成盘根错节的关系。举凡高校研究基地评审，硕博导评选，研究生培养指标的分配，针对教师的各类考核、评奖与评优，人才的录用与引进等，帽子、论文、项目、奖项、职称等往往成为主要参考依据。在高校林林总总的各类学术评价中，以学术论著为代表的学术产出是评价的基石，在某种程度上充当了学术评价中一般等价物的功能。因此，在高等教育领域的学术评价中，唯论文是“五唯”的核心病灶所在，“破五唯”的成功落实与否在很大程度上取决于对唯论文倾向的纠正。当然，“唯论文”是一种表达上的修辞，其所指不局限于学术论文这一种形式，而是指对学术论文、发期刊影响因子或层级、论文引用次数和引用率、学术著作及其出版社等级、发明专利和产品等一系列学术作品外在特征进行计量，并据此做出评价结果的一种学术评价方式。

我国高校以学术论著作为学术评价主要依据的历史并不久远，在引入论著评价之前，高校学术评价长期存在着标准不清晰、程序不透明的弊病。以高校教师最为关切的职称评定为例，在论著评价尚未普及的20世纪80至90年代，评审委员会的合议是职称评定的主要方式，在缺乏其他相对客观的评价标准的情况下，教师的年资往往成为评价的重要参考。在这种制度设计中，职称评定往往沦为部分学术权威把持下以论资排辈为主要工具，辅之以对教学科研业绩定性考察的学术资源分配制

度，带有计划经济体制下的平均主义烙印，也存在大量寻租空间。这既不利于青年人才的脱颖而出，也不利于激发教师的学术活力与科研潜力。1990年，南京大学首先将学术论著尤其是SCI论文引入高校学术评价，此举在当时被视为破除学术评价中论资排辈与平均主义弊病的创举，也极大地激发了教师的科研活力，肇始者南京大学的SCI发文数量也连续多年位居全国第一。然而以学术论著作为学术评价主要参照物的做法风靡一时并最终成为众矢之的，却并非出自政府的倡导。罗燕认为以大学排行榜为代表的第三方独立评估是“唯论文”的主要推手，固然能够部分解释其产生的社会机制，但也很难说明高校主动采用论著评价以及政府部门默许这一做法的内部动力。这可能与部分高校的示范效应和学术论著本身所具备的特征有关：一方面，论文数量、发表期刊的层级等指标具有客观性与刚性的特点，在具有淘汰赛性质的各类学术评价中可以对评价对象进行横向比较，通过“点数字”的方式很容易区分高下；另一方面，数量与层级是一种外在于评价者的客观标准，以论文数量与期刊层级进行评价可以降低评价过程的主观性以及人为操作空间，在历史上也曾充当了

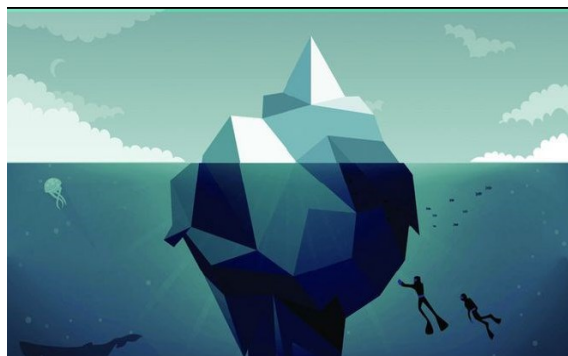


搅局者角色，破除当时以年资和圈子关系为基础结成的牢固网络，为青年教师发展提供了更好的学术环境。然而量化评价有其固有缺陷，它在形式上以客观和可比较的数字化形式把学术评价简化成了一种点数字游戏，几乎可以脱离论著的内容本身来实施评价，本质上沦为高校科层制的管理工具，而非基于论文实际学术贡献的学术评价，因此很难识别论文真正的学术贡献。

**“唯论文”倾向在高校学术评价中大体表现在三个方面。其一是除在职称和人才称号评定、教师绩效分配、年终评优评奖等传统学术评价领域之外，论文作为一个好用的评价工具与计量指标已扩散至几乎所有评价活动，如研究生培养指标的分配，研究生的毕业条件、评优评奖、出国入党资格的评定等，其学术硬通货的角色愈发突出。其二是在各类具体的学术评价中，即使制定了多元评价指标，论文在其中仍然起着一锤定音的作用，典型的如高校教师职称评定中，对论著总量、发表论文的期刊层级及其对应数量、不同层级期刊论文的数量换算比例等均有细致规定，而对教学工作、社会服务工作等，往往只进行数量的底线约束或描述性规定，在评价过程中论著起着决定性作用。其三是当论文作为评价工具时，往往以论文数量、发表期刊层次等外在形式作为评价标准，而忽视对论著内容本身的分析并在此基础上对其研究水平与研究贡献做出判断，体现出强烈的量化评价取向。以层级高低、数量多寡取代研究水平与贡献评价是“唯论文”倾向的底层逻辑。**

“破五唯”自上而下的议程设置使高校以执行者的角色出台各自的“破五唯”措施。然而，倘若未能事先明确可行的替代方案，会使高校学术评价面临失去依归的风险。一个可行的替代方案不仅应明确学术论著在学术评价中的地位，更应为学术论著评价设计评价标准和操作规程。

**首先，破“唯论文”倾向并非在学术评价中将学术论著弃置不用，学术论著作为学术场域一般等价物和学术评价核心标的物的地位是科学共同体一致认可的，它不仅是科学发现的发布方式和学术交流的平台，也是学术贡献的认可方式和学者个人学术声誉的来源。**所谓破五唯，一是要明确学术论著作为评价标准的适用范围，作为一种学术评价标准，学术论著应运用在对学术单位、学术人员的学术业绩与学术水平的评鉴中，将论著评价撤离与学术活动关系不甚密切的诸如研究生招生名额分配、研究生评奖评优、研究生答辩资格审查等领域，使之回归学术贡献与学术水平评价本位，避免



将论著评价作为各种甄别的政策工具。二是要纠正学术论著在评价标准中一家独大，乃至成为评价结果唯一影响变量的倾向，例如在学科评估中除学术论著体现的科学研究水平和贡献外，应重视学科在优质人才培养、经济增长、社会发展等方面的贡献；在职称评定中要重视教师在实现立德树人这一高校根本任务中的贡献，在对教师教学业绩的评价中淡化基础教学量的刚性要求，重点考察教师在人才培养制度改革、课堂教学方式革新、课堂教学效果提升等方面的措施与贡献。三是在学术论著评价中要破除以数量、刊物（出版社）层级等外在形式作为评价依据的KPI主义，抛弃精确量化的评价方式，把论著所体现的贡献和研究水平作为评价的主要依据，通过对论著的内容分析，考察学术论著对于人类知识增长、前沿科技突破、重要技术创新以及其他关乎国计民生、社会治理、文化创新等问题的认识与解决的实质性贡献。

**要破除论文数量迷局，推行代表作制度不失为一个良好策略。**论著代表作在作为评价材料提交评审时，已然经过期刊社（出版社）及其评审人、作者本人以及时间维度上的学术影响力等多重检验，基于代表作的学术评价可以使评审过程更多集中在对论著内容所代表的学术水平与学术贡献进行评价，免于论著发表数目等变量的干扰，因而代表作评价能够较为准确地识别评价对象的学术水平与学术贡献。推行代表作评价制度，可以使高校教师从论著生产的烦琐考核中解脱出来，精心地选择研究问题和组织研究，力图做出最有价值、最有代表性的研究，从而达致引导教师从事高质量研究的目的。尽管实施代表作制度难免导致总产出的降低，但这是提高研究质量与贡献的可承受代价。

**落实代表作制度，需要进行相配套的绩效考核制度与绩效激励制度改革。**对教师的绩效考核不宜过频，淡化年度考核的终结性评价性质，允许教师以研究内容和研究过程描述作为年度考核材料，为教师投入周期较长的高质量研究提供相对宽松的环境。教师工资、绩效激励不宜与论著产出数量密切挂钩，应以普惠式的生活保障为主要目的，使教师不致为挣工分而片面追求论著数量。

**破除“唯论文”倾向，还应改进评审制度。**当前主流的同行评审制度，无论是会议评审、无记名投票的审议制度，还是匿名通信评审制度，其决策过程都颇为隐晦，当事人与旁观者无法判定评审者是基于怎样的考量做出评议决策，也无从监督评审者是否基于论著学术价值做出自己的判断。对此，可否考虑将关涉教师发展核心利益的学术评价（如职称评定）从封闭式会议评审改为开放式评审，借鉴学位论文答辩的某些形式，由评审者和被评审者就论著的具体内容、学术贡献和学术水平进行公开的陈述、辩论与评价。在通信评审环节，有条件的高校或学科，亦可考虑引入国内国外双重通信评审。

