



中国教育科研参考

2020年第14期
总第(480)期

中国高等教育学会编

2020年7月30日

目 录

欧盟“地平线 2020 计划”的监测与评价:理论、方法及启示.....	曲瑛德 赵 勇 (02)
“双一流”建设背景下大学评价的改进方向.....	宫新栋 杨 平 (10)
我国“双一流”建设实施绩效评价面临的困境与行动路径.....	杜 瑛 (15)
世界一流大学五大评价指标体系的比较、改进及其启示.....	赵国栋 马瑞敏 (22)
中国特色世界一流大学建设成效评价体系理论建构与实践验证.....	冯用军 赵 雪 (29)

编者的话:“双一流”建设是我国为推进高等教育事业跨越式发展、建设高等教育强国而实施的一项战略性工程。为了推进这一工程,中央和地方政府出台一系列文件、投入大批建设经费,“双一流”建设高校在政策和经费支持下纷纷加大学科建设力度、提升科学研究水平,创新学校管理理念。目前,“双一流”建设的评估验收和调整期限即将到来,这些高校的“双一流”建设成效如何、如何实施有效评价,成为高等教育管理部门和各大高校关注的热点问题。为了深入认识这一问题,本刊以“‘双一流’建设评价”为选题,集中选编若干文章,供读者参阅。

主编:王小梅 本期执行主编:王者鹤 责任编辑:薛欣欣
地址:北京市海淀区学院路35号世宁大厦二层《中国高教研究》编辑部
邮编:100191 电话:(010)82289809
电子信箱:gaoyanbianjibu@163.com
网址:www.hie.edu.cn(中国高等教育学会——学术观点栏目)

欧盟“地平线2020计划” 的监测与评价：理论、方法及启示

曲瑛德 赵 勇

一、引言

“地平线2020计划”于2014年开始实施，是欧盟历史上规模最大的研发创新计划，其前身可追溯至1984年欧盟提出的“研究、技术开发及示范框架计划”（简称“框架计划”），该计划旨在通过欧盟层面的资源整合，加强研发和创新能力，提升全球科技竞争力，从而确保其在全球市场中的优势地位。至2013年，欧盟框架计划共实施了7轮，累计预算1110.27亿欧元。2011年，为消除欧洲债务危机的影响，提高科技创新效率，欧盟委员会提出整合框架计划（FP）、竞争与创新计划（CIP）和创新与技术研究院（EIT）三大科技创新计划，决定在第7框架计划（FP7）结束后实施“地平线2020计划”，总预算将近800亿欧元。2013年，欧洲议会和欧盟理事会联合发布《实施地平线2020：研究与创新框架计划（2014—2020年）》（第1291/2013号）规章，制定了“地平线2020计划”监测与评价的相关条款，要求欧盟委员会基于定量和定性证据对计划执行情况进行年度动态监测，并依规开展中期评价（2017年底完成）和事后评价（2023年底完成）。

计划评价（Project Evaluation）是针对计划实施的效果、效率和影响等方面开展的以证据为基础的评测与研判过程。然而，如何评价计划实施所产生的科技、经济、社会等影响一直是本领域的方法论难点，例如影响的时滞、归因、贡献、不确定性、风险等都是计划评价中的复杂性问题。“地平线2020计划”基于项目理论（Programme Theory）方法和最新的信息技术手段，开发了一整套的指标系统，构建了有效的监测与评价体系，从而创新了计划评价的理论与实践，同时满足了欧洲议会和欧盟理事会的政策需求。从计划实施的目标和举

措上看，我国正在开展的“双一流”建设项目与欧盟“地平线2020计划”存在一定的相似性。“双一流”建设成效评价本质上属于计划评价的范畴。因此，本研究以欧盟“地平线2020计划”为研究对象，重点对年度监测报告、中期评价报告等相关资料进行系统调研，揭示其监测和评价体系的内部原理、基本特征及相互关系，进而从评价思路、评价内容、评价方法、评价组织等多个维度总结经验和启示，为我国“双一流”建设成效评价的体系设计与组织实施提供参考。

二、文献综述

（一）计划评价的挑战与应对

由于知识生产和传播过程的复杂性，针对研发创新计划的评价面临诸多方法论上的挑战。首先，计划评价需要基于证据对计划实施的效果和效率进行评测，也需要对计划成果的影响做出判断，尤其是还应立足更广阔、更深远的视角，对计划成果的长远影响和可能影响作出研判。其次，计划评价不仅是评测计划实施带来的变化程度，还要考虑变化背后的因果关系（Why），其超越了动态监测与合规审计的范畴。动态监测重点关注的是计划实施的状况是什么（What），如发生了什么变化、计划产出了什么结果等。合规审计则重点关注计划如何运行（How），如内部机制如何运作、资源如何被使用等。此外，研发创新计划自身就是一个复杂系统，涉及目标、组织方式、干预手段、参与群体等多要素和全流程，这也给评价带来了诸多难题。

国际上，部分研发资助和管理机构建立了一些监测和评价体系用于洞察计划项目实施所产生的综合影响。Guthrie等曾对全球14个科研评价体系进行了系统调研，并根据评价目标将体系划分为宣导型、问责型、学习型和资源分配型四类，发现问

责和分配型评价体系的方法选择并不适用于学习型评价体系。前者主要基于定量方法路径，强调透明性和可比性，而后者主要基于定性方法路径，注重综合性和灵活性，并不需要比较。

目前，研发创新计划的监测与评价体系仍以问责型和资源分配型为主，主要依据计划干预的逻辑，将计划的相关信息表征为投入、产出、结果、影响的相关指标，以定量分析为主展开评价活动。传统的定量评价方法主要是基于文献计量数据（如论文、专利、奖励等），利用产出和引用指标来测度计划干预对科技创新的影响。这种方法基于“卓越科研促进社会发展”的前提假设，因此其反映的只是研发计划对科学发现和技术发明的影响。

近年来，随着研发计划投入力度的不断加大，对其经济和社会影响评价的需求也越来越凸显。在经济影响评价上，企业的就业和经营数据已经被用于生产力和经济增长的影响估算模型中。但是，对研发创新产出的统计分析却仍局限于专利，并没有拓展至其他知识产权类型，如商标或标准等，同时也缺少对科技成果市场转化的数据跟踪。

目前，芬兰国家商务促进局在其监测系统建设中对研发的经济影响测度开展了指标设计的积极探索。在社会影响评价上，学界提出基于社交媒体和政策文本等数据，利用替代计量方法（Altmetrics）对研发计划的社会影响进行评价，但也有学者认为替代计量方法测度的是研发产出的传播情况而非社会影响。另外，也有一些其他的方法路径尝试，如在欧盟第七框架资助下，荷兰皇家艺术和科学学院主导的SIAMPI项目通过分析科学与社会之间的互动关系，从社会影响的关键因素指标切入，对研发资助计划的社会影响评价方法进行了研究。

整体上看，定量和定性方法相结合是研发创新计划评价的发展趋势。然而，数据仍然是当前计划评价面临的最大挑战。尤其是在大数据环境下，单维度数据不利于计划评价的高质量实施，只有在开放和互操作的数据基础平台上，评价指标方能实现其潜力。

（二）以理论为基础的计划评价

项目理论是指通过多种途径构建因果模型，将计划的投入、干预活动与已有或预期结果相链接，并基于该模型开展评价。项目理论的核心方法论包括两个部分，一是变革理论（Theory of Change）方法，重点分析计划通过哪些活动来实现短期、中期或长期目标，以及计划背后可检验或可衡量的关键性假设；二是行动理论（Theory of Action）方法，重点阐释计划干预的机制，即计划如何来“激活”变革。项目理论也有其他近似的术语表述，如理论驱动评价、项目逻辑、干预逻辑、项目理论驱动的评价科学等。单线性因果模型是项目理论的常用逻辑模型，主要包括投入、行动、产出、结果和影响五类变量。也有学者探究了利用非线性模型来指导计划评价，并将“复杂性”的概念尝试引入到模型的构建中。目前，欧盟委员会在《良好规章制度准则》中已经将项目理论作为计划评价的基本准则，并广泛地应用于其主导的研发创新计划评价之中。

计划评价从项目理论到指标系统还需要一系列的具体方法支撑。首先，影响路径分析（Impact Pathways Analysis）是在项目理论指导下构建指标系统的首要环节，其目标是确定不同阶段（短期、中期、长期）计划结果产生综合影响的具体路径。Rogers认为在复杂背景下基于项目理论开展计划评价时，需要尽可能覆盖计划干预的所有影响路径，在没有深入定性分析和参与式评估的基础上，不能基于量化指标体系利用逻辑模型开展评价。其次，在定性分析之后，评价的另一个难点是确定计划的附加值，也就是阐释影响的产生是计划实施的结果，而非其他因素所致。计划评价一般采用反事实分析（Counterfactual Analysis）方法来解决这一问题，其目标是确定在没有计划干预下的可能情况，并将其与实际情况进行比较。反事实分析需要微观经济学研究方法和细粒度数据的支撑。再次，将计划的实际效果与预期目标进行比较时，需要设定好基线（Baseline）。因此，具体指标的选择还需要考虑到数据的完整性、连贯性与一致性。

最后，在大数据时代，人们可获取的数据在不断增多，海量、多源、异质数据的融合与利用需求也在不断催生新的技术手段。因此，吸纳最新的大数据分析技术，不断提高指标数据采集和分析的智能化程度也是计划评价方法的发展趋势。

三、“地平线2020计划”监测体系的主体框架与核心内容

监测（Monitoring）是对计划活动及其影响随时间推移所产生的代表性（指标）数据进行连续和系统地测定，其有助于识别和解决计划实施过程中出现的问题，同时为后续评价和影响评估提供事实数据。根据第1291/2013号规章要求，欧盟委员会对“地平线2020计划”的具体监测对象包括：三大战略优先领域（卓越科学、产业引导和社会挑战）及其子计划、四个单列资助计划（倡导卓越和扩大参与、社会参与科学和科学服务社会、创新与技术研究院、欧洲原子能共同体研究与培训），以及14个交叉问题。[实现欧洲研究区域、扩大参与、中小企业参与、社会科学和人文科学、科学与社会（负责任的研究与创新）、性别、国际合作、可持续发展与气候变化、科学发现与市场应用的衔接、数字化议程、私营部门参与、PPP和P2P资助、交流与传播、独立专家的参与方式]“地平线2020计划”监测体系主要包括利益相关者年度调查、“地平线2020计划”指标系统、信息技术支持系统、年度监测报告四个方面，其核心思想是在信息技术的支撑下，通过对定量证据（以指标数值结果呈现）和定性证据（以利益相关者调查结果呈现）的采集与分析，形成年度动态监测报告制度，以保证计划的有序开展和良性运行。

（一）利益相关者年度调查

利益相关者年度调查是动态监测体系的一项重要支撑内容，也为“地平线2020计划”中期评价结果解读提供了语境化的定性辅助资料。该调查由欧洲研究委员会的国家联络点（NCPs）负责实施，其目的是针对计划的参与和实施情况广泛搜集意见、评论和建议，重点聚焦计划对利益相关者的吸引力、计划目标与研究创新需求及欧洲2020战略的

相关性、计划与欧盟其他资助的一致性、计划实施的附加值。调查问卷通过邮件方式发送至联络点覆盖的欧盟成员国和计划参与国。根据计划实施的进度，历年利益相关者调查问卷的问题设置有所不同。2015年首次调查设计了30项多选题和1项开放性问题，集中在以下4个方面：一是“地平线2020计划”的政策工具作用，涉及区域和国家层面的一般性政治话题，以及与其他欧盟计划活动的互补性；二是开放科学，涉及科学对整个社会的促进作用，包括企业、性别和伦理议题；三是开放创新，涉及私营部门的贡献，包括公私合作，以及创造就业和增长方面面临的艰巨挑战；四是向世界开放，涉及国际化维度，包括欧洲的全球角色。2016年第二次调查则设计了34项多选题和1项开放题，主要集中在：一是吸引力，涉及计划项目的立项、科学与市场的衔接、第三方国家参与、中小企业参与；二是交叉问题，涉及欧洲研究区域、性别、社会科学和人文科学、负责任的研究、可持续性；三是欧盟框架计划的附加值，涉及单一欧共体成员国难以有效实施的目标和行动。

（二）“地平线2020计划”指标系统

应用指标来洞察“地平线2020计划”的结果和影响是动态监测体系的一个核心内容。指标系统的建立也为评价体系的有效运行奠定了坚实的数据基础。欧盟“地平线2020计划”确定了五项原则用于开发目标明确的指标系统。一是靠近原则（Proximity），采用VAT税号、研究人员ID，资助者ID等唯一标识符来确定计划评价中涉及的具体研究人员和企业。二是归因原则（Attribution），超越出版物、专利等传统指标，全面反映因计划实施而产生的科学、经济和社会影响。三是可追溯原则（Traceability），基于外部公共与私营数据库开发标准化数据；使用项目官员、评价者、审核者的其他相关数据源。四是整体原则（Holism），将计划进展作为一个整体来体现，关注不同计划行动的长期共同目标和相互联系。五是稳定性原则（Stability），建立信息系统，确保与早期框架计划的延续性和可比性；增强数据质量，在监测体

系中尝试不同的数据收集和分析方法。“地平线2020计划”关键绩效指标系统主要涉及卓越科学、产业引导和社会挑战三大战略优先领域，倡导卓越和扩大参与、社会参与科学和科学服务社会、创新与技术研究院、欧洲原子能共同体研究与培训4个单列资助计划，以及14个交叉问题。指标数据主要来自公共研究数据库（CORDA），其收集了从欧盟第五框架计划到“地平线2020计划”所有研发创新项目的相关数据，但仅限于欧盟委员会内部使用。

（三）信息技术支持系统

在前几轮框架计划实施过程中，欧盟委员会陆续建立了一系列的信息系统平台，用于支撑研发创新项目信息的管理与利用。“地平线2020计划”将已有信息系统资源进行了关联，实现了数据之间的融通，为其监测体系的有效运转提供了坚实的信息技术系统保障。目前在运行的主要信息系统包括：计划书提交和评价系统（SEP）、合同与计划管理系统（CPM）、资助管理系统（SYGMA）、科研绩效与影响报告工具（RESPIR/SESAM）等。公共研究数据库（CORDA）就是在已有信息系统基础上对数据进行了结构化处理，为“地平线2020计划”指标系统提供了数据支持。此外，针对特定用途，欧盟委员会还设立3个信息平台建设项目。一是研究与发展信息服务平台（CORDIS），其收集了欧盟第四框架计划以来所有研发创新项目的基本信息，进展、阶段性成果和总结报告等，并将项目信息与项目产出的出版物和专利信息进行了数据关联。二是开放获取基础设施（OpenAIRE），其为计划项目的研究成果提供了开放获取服务，促进了欧洲开放科学发展。三是欧盟开放数据门户（The EU Open Data Portal），提供了大量欧盟委员会及其他欧盟机构的开放数据资源。

（四）年度监测报告

“地平线2020计划”的年度监测报告由欧盟委员会的研究与创新总署负责发布，是对计划监测结果的整体呈现，也是“地平线2020计划”中期评价的主要参考资料。年度监测报告主要涉及以下5个部分。一是计划立项情况，包括立项数量、资助

金额、申报数量、成功率、申报与立项机构类型、欧盟成员国的参与和绩效。二是计划整体运行情况，包括新获批项目者、项目时间、精简程序、与其他资助计划的协同、计划书评审的质量评价、纠正、伦理、安全审查程序。三是具体计划运行情况，基于“地平线2020计划”关键绩效指标系统，重点监测三大战略优先领域、四个单列计划和14个交叉问题领域的指标变化。四是资助计划的案例分析，涉及开放创新、开放科学和向世界开放三个领域案例。五是利益相关群体年度调查结果。

四、“地平线2020计划”评价体系的逻辑模型与分析方法

评价（Evaluation）是立足更广的视角对计划实施的成效进行审视，重点关注计划实施是否带来变化，是否向既定目标迈进。具体来说，在政策目标实现过程中，计划实施带来什么变化，变化的原因是什么，变化的程度如何。欧盟“地平线2020计划”的评价体系主要依据项目理论而构建，包括逻辑模型、影响路径分析和评价方法3个核心部分。其评价数据来源不仅包括年度监测报告的指标内容，还包括欧盟统计局（Eurostat）和经合组织（OECD）数据，欧盟委员会的专题评价结果（如交叉问题）、专项研究结果（如基于Scopus数据的出版物和网络分析、欧盟框架计划附加值和经济影响评价），以及欧盟其他机构提供的数据。

（一）逻辑模型

“地平线2020计划”评价采用了项目理论中的单线性因果关系逻辑模型，即将计划项目信息结构化为“目标-投入-行动-产出-结果-影响”六大要素，同时提出相关性（relevance）、效果（effectiveness）、效率（efficiency）、一致性（coherent）和附加值（added value）五个评价问题。（见图1）其中，相关性评价的是“地平线2020计划”与既定目标的符合程度，需要回答计划行动是否满足现实的需求。效果评价的是“地平线2020计划”是否达到或推进了其既定目标，需要回答现实发生的变化多大程度是计划实施的结果。效率评价的是资源投入与计划项目产出之间的关

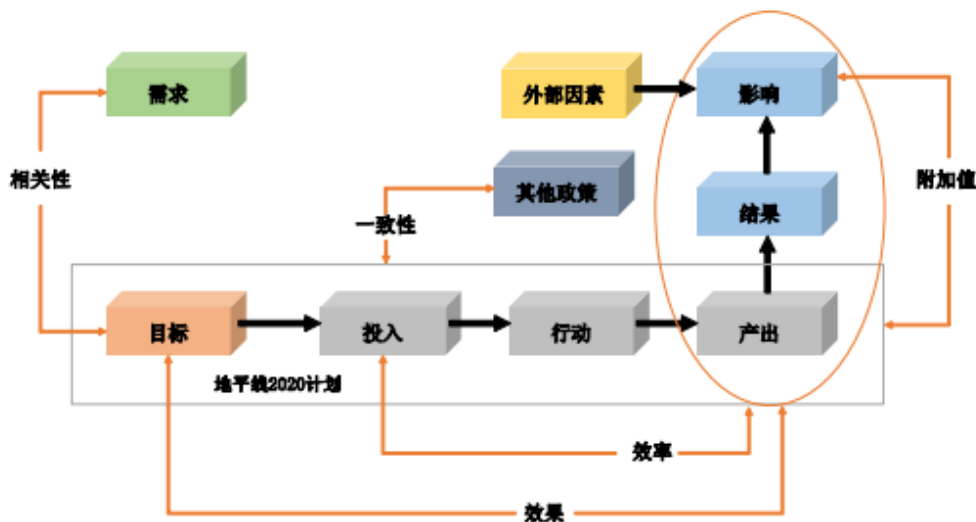


图1 欧盟“地平线2020计划”评价的逻辑模型与评价问题示意

资料来源：European Commission. Better regulation “Toolbox”, 2017: 336。

系，重点关注计划项目的成本与效益，需要回答计划的成本投入在多大程度上是有效的。一致性评价的是“地平线2020计划”内部项目之间、以及与欧盟其他资助计划之间的协调性。附加值评价的是“地平线2020计划”在区域或国家层面实施的干预活动所产生的价值。

（二）影响路径分析

“地平线2020计划”评价体系按照关键影响路径分析结果将年度监测内容进行了重新归置。关键影响路径分析主要关注计划实施产生的短期、中期和长期的最典型变化，并不局限于“地平线2020计划”管理和运行的监测数据。通过计划干预的影响路径分析，“地平线2020计划”的影响被具体划分为科学影响、创新/经济影响和社会影响三大类。其中，科学影响的关键路径包括：欧盟世界一流的卓越科学、更好的跨界和跨部门合作及研发工作整合、欧盟科学新领域的兴起。创新/经济影响的关键路径包括：更好的欧洲企业创新能力、欧盟技术引领和竞争力增强（包括中小企业、初创企业和大型企业）、经济创新的扩散（包括中小企业）创造就业、增长和投资。社会影响的关键路径包括：研发更好地应对社会挑战、增强欧盟在应对社会挑战上的全球示范作用、更适用于社会的科学与创新方案。

（三）评价方法

“地平线2020计划”中期评价主要是以欧盟第七框架计划事后评价结果为基线，针对相关性、效率、效果、一致性和附加值5个方面问题开展整体评价，其评价方法综合采用了文献计量学、经济学、统计学、管理学等多学科的研究方法，同时吸纳了文本挖掘等新的大数据分析技术手段。

1. 相关性评价方法。相关性评价主要采用文本挖掘和词云可视化方法对“地平线2020计划”信息抽取的关键词与政策文件、社交媒体、专利和出版物抽取的关键词进行匹配度比较分析，进而评价“地平线2020计划”与宏观战略需求和现实问题之间的符合度。在“地平线2020计划”中期评价中，文本挖掘的计划信息包括：欧盟委员会实施“地平线2020计划”的决议（第2013/743/EU号）、历年工作计划等；政策文件信息包括：欧洲2020战略、欧盟理事会“巴黎协议书”、联合国2030可持续发展议程、世界健康组织报告（2015）、G7领导人宣言（2016）等；社交媒体信息包括：推特、脸书、维基百科等；出版物信息主要来自谷歌学术。

2. 效率评价方法。效率评价主要基于成本效益分析方法，重点考察计划投入（如资源、预算、计划书评选）与产出之间的关系。“地平线2020计划”中期评价主要关注了项目管理和执行过程中的效率，包括计划的行政成本、精简程序的受益（如计划书撰写成本）、欧盟委员会的服务（如计划的行政和运行成本），以及资助分配可能产生的目标偏差（如计划参与国家的类型是否符合向世界开放的目标）。其中，计划书撰写成本是欧盟委员会在已有相关研究基础上进行的估算。计划参与国的合作模式则采用了计划参与的网络分析和计划产出的

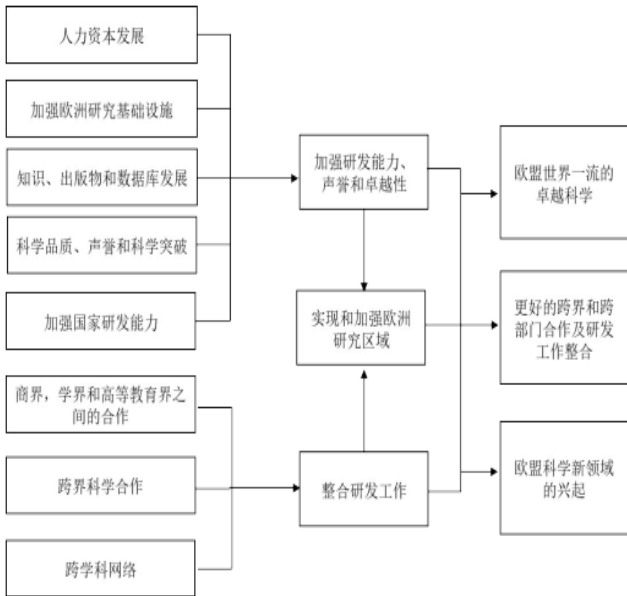


图2 “地平线2020计划”科学影响的路径分析

资料来源:European Commission. Better regulation “Toolbox”, 2017: 107。

出版物分析（文献计量学的共现分析方法）等两种方法。

3. 效果评价方法。效果评价重点考察“地平线2020计划”的目标实现情况。“地平线2020计划”中期评价在效果评价中主要分析了计划产生的科学影响、创新/经济影响和社会影响。在科学影响评价上，中期评价重点关注了“加强研发能力、声誉、卓越性”和“整合研发工作”两个子路径的8个具体策略。（见图2）多数具体策略的评价都挑选了关键绩效指标系统中的具体指标数据，如人力资本发展的测度指标包括：“玛丽居里”行动中跨国流通研究人员数量、学界与非学界之间流动研究人员数量、欧洲科研人员网络（EURAXESS）发布的年度研究职位数量。此外，跨学科性评价主要是基于期刊论文的引文数据，将论文引用的外学科文献比例作为其跨学科程度的判断依据。

在创新和经济影响评价上，“地平线2020计划”中期评价重点关注了7个社会挑战，以及“加强科学与社会之间联系”和“加强研发作为政策制定的基础”子路径下的7个具体策略。（见图4）在7个社会挑战方面，社会影响评价挑选的关键绩效指标有：同行评议期刊出版物，专利申请和授权，投放市场的新产品、新工艺和新方法等。此外，中期评价还采用了“地平线2020计划”合作者调查（PPMI）结果数据作为支撑，并对典型计划项目的社会影响进行案例分析。

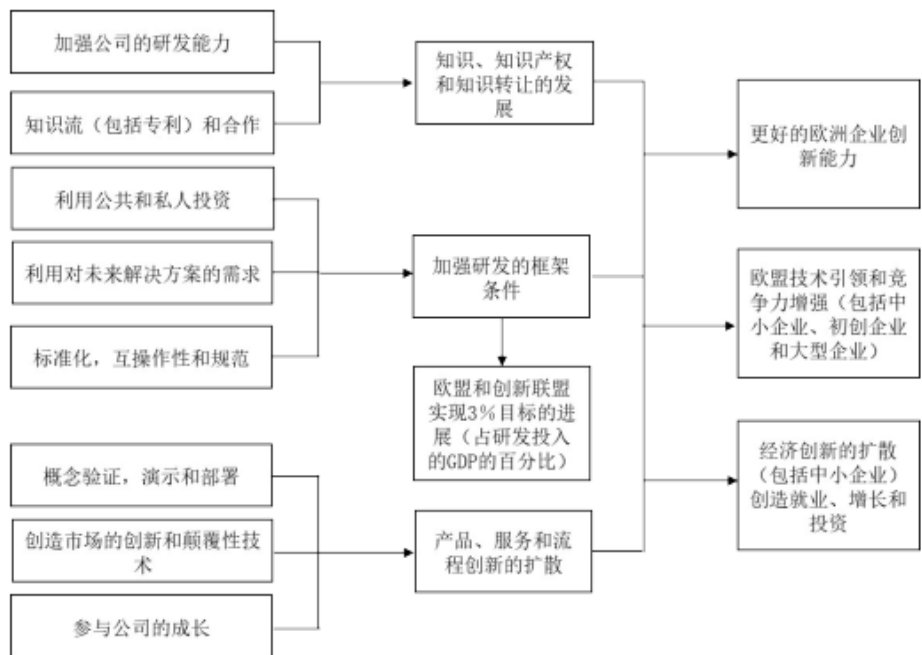


图3 “地平线2020计划”创新和经济影响的路径分析

资料来源:European Commission. Interim Evaluation of Horizon 2020, 2017: 130。

在创新和经济影响评价上，“地平线2020计划”中期评价重点关注了“知识、知识产权和知识转让的发展”“加强研发的框架条件”“产品、服务和流程创新的扩散”三条子路径的8个具体策略。（见图3）除将年度监测报告中的指标内容按照具体策略重新归置外，经济影响评价还采用了宏观联接计量经济模型（NEMESIS）。

在社会影响评价上，“地平线2020计划”中期评价重点关注了7个社会挑战，以及“加强科学与社会之间联系”和“加强研发作为政策制定的基础”子路径下的7个具体策略。（见图4）在7个社会挑战方面，社会影响评价挑选的关键绩效指标有：同行评议期刊出版物，专利申请和授权，投放市场的新产品、新工艺和新方法等。此外，中期评价还采用了“地平线2020计划”合作者调查（PPMI）结果数据作为支撑，并对典型计划项目的社会影响进行案例分析。

4. 一致性评价方法。一致性评价主要是在计划的影响评价基础上进行，重点考察“地平线2020计划”内部项目之间、以及与欧盟其他政策或干预措施之间的协同性。“地平线2020计划”中期评价

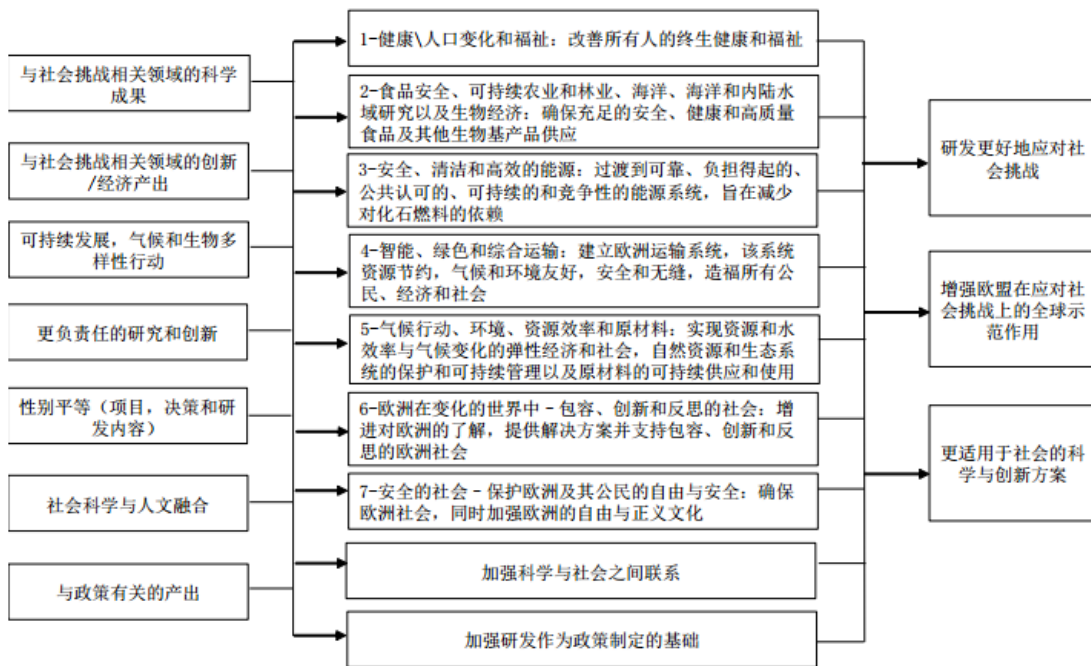


图4 “地平线2020计划”社会影响的路径分析

资料来源: European Commission. Better regulation “Toolbox”, 2017:160。

果、效率评价, 拓展到计划实施对科学、经济和社会发展的影响; 评价方法向定量与定性相结合、指标细化评价方向发展, 加强了对文本挖掘、文献计量、专利分析、社会网络分析等新技术的吸收;

中, 内部项目间的协同性评价主要采用了主题评价的交叉分析方法, 将计划与服务的目标群体相互对应, 对不同计划项目之间的研究人员数量和年龄、项目数量和预算等指标进行比较。同时, 以利益相关者调查结果定性辅助资料展开分析。与其他政策或计划间的协同性评价则重点关注了“地平线2020计划”与欧盟结构与投资基金 (ESIF)、共同农业政策 (CAP)、大型和中小型企业竞争力计划 (COSME), 以及欧盟伊拉斯莫斯计划和生命计划之间的相互关系。

5. 附加值评价方法。“地平线2020计划”中期评价对附加值的评价主要采用了“欧盟框架计划附加值和经济影响评价”专项研究结果, 其以欧盟第七框架计划事后评价结果为基线, 利用了经济学的断点回归和倾向匹配方法, 对“地平线2020计划”进行了反事实分析。同时, 也采用了案例分析方法研究了典型计划项目的附加值情况。

五、对我国“双一流”建设成效评价的启示

欧盟“地平线2020计划”监测与评价体系是对现有计划评价的理论与实践创新, 其评价思路由计划项目个体评价向整体计划评价转变, 将静态评价与动态监测相结合, 注重计划“制定-实施-效果”全过程的分析; 评价内容由政策和项目实施效

评价组织呈现多元化, 强调政府、学界和业界之间的对话与合作。当前, 我国高校评价体系存在诸多问题, 如评价理论薄弱分散, 将高校从其依存的社会情境中抽离, 未能聚焦我国的经济社会发展和国家重大战略发展的现实需求; 评价方法相对简单, 多以统计方法为主; 评价内容过于单一, 未能全面反映高校在人才培养、社会服务等方面的贡献。

因此, 从方法论层面总结欧盟“地平线2020计划”监测与评价体系的经验, 将对我国即将开展的“双一流”建设成效评价具有重要的参考价值。

(一) 采用系统性思维, 形成整体性项目评价思路

创建世界一流大学和一流学科是一项长期性和系统性的建设任务, 涉及加强党对高校的领导、完善内部治理结构、建设一流师资队伍、培养拔尖创新人才、提升科学研究水平、传承创新优秀文化、推进成果转化等多项建设任务和改革任务, 任务之间存在相互关联和相互作用。系统性思维强调整体性, 强调从系统的结构去认识系统的整体功能, 并从中寻找系统最优结构, 获得最佳系统功能。因此, “双一流”建设应立足项目的整体评价思路, 借鉴项目理论提出的逻辑模型, 确定评价的核心问题 (如达成度、贡献度、引领度、支撑度和

满意度），构建政府管理部门、“双一流”建设高校、专家委员会、第三方评价机构、其他利益相关群体多方参与的评价网络，同时将“双一流”建设的年度进展报告、中期评估报告、动态监测报告等内容作为成效评价的主要数据来源。

（二）超越传统文献计量标准，构建多维度指标体系

欧盟“地平线2020计划”评价超越了传统的文献计量学范式，利用多维度指标考察了研发计划在科学进步、经济发展和社会影响方面的贡献。当前，我国高校的评价体系存在唯论文、唯帽子、唯职称、唯学历、唯奖项的顽瘴痼疾。从根本上解决评价的指挥棒问题，就要求“双一流”建设摒弃以论文、奖项为主要评价标准的做法，构建创新、多元、综合的评价维度来反映建设成效。在指标体系设计时，需要考虑学科的差异性，按照自然科学、工程科学、人文科学、社会科学大类进行构建。同时，综合定量指标与定性指标，推行“代表作”制度，更多地选择标志性成果的代表性指标，强调成果的“特色、服务、贡献、引领”特征。此外，加强对交叉学科建设情况的指标监测，适当引入当前较为成熟的交叉学科测量指标。

（三）面向国家重大战略需求，突出综合影响评价导向

“双一流”建设是全球知识经济背景下，国家提升科技竞争力和综合实力的重大战略部署，其要求“双一流”建设高校要面向国家重大战略需求，面向经济社会主战场，面向世界科技发展前沿，突出建设的质量效益、社会贡献度和国际影响力。因此，我国“双一流”建设可以在项目理论指导下，以综合影响评价为导向，对关键影响路径开展分析，在教育、科技、经济、社会等方面确定影响的主路径和子路径。同时，关键影响路径的确定可以吸收欧盟经验和理念，立足中国特色，具备全球视野。在此基础上，合理使用动态监测指标数据、官方权威统计数据、第三方评估数据作为成效评价的综合数据来源。此外，综合影响评价中也应

突出体现学科的建设特色、高校的区域特色对国家重大战略需求的贡献，引导高校把握自身特色，实现差异化发展，更好地服务于国家、行业、区域发展等。

（四）吸收多学科方法与技术，提高评价精准化水平

综合定量和定性方法是当前计划评价的发展趋势。“双一流”建设成效评价方法应吸收多学科方法的特色，针对不同评价对象采用多样性的评价方法。同时，将数据挖掘、自然语言处理、可视化图谱等大数据分析方法和应用于评价之中，增强评价结果的深度和准确性，也减轻被评价方在评价过程中的负担。

为此，“双一流”建设主管部门需要成立专项研究课题组，针对评价中的方法论难点展开研究，形成科学化的评价方法指南，指导“双一流”建设高校开展终期的成效评价活动。

（五）加强信息系统平台建设，打造智能化评价体系

智能化的评价系统平台要以信息科学的方法为基础，计算机科学的技术为保障，为管理部门和受评高校和学科提供互动的平台，同时也要以“节约成本、减轻负担”为原则，提高系统平台的运行效率。另外，数据是“双一流”建设成效评价的基础，数据的质量决定着评价结果的准确性。所以，“双一流”建设成效评价也应加强对数据的组织和管理的工作，利用系统平台完善评价数据的收集、整理与再利用工作，通过关联数据等最新技术提高评价数据的语义化程度，为“双一流”建设的科学化评价体系建设提供支撑。

（原文刊载于《中国高教研究》2020年第1期）

（曲瑛德，中国农业大学发展规划处处长，北京 100193；赵勇，中国农业大学情报研究中心研究馆员，发展规划处副处长，北京 100193）

“双一流”建设背景下大学评价的改进方向

宫新栋 杨 平

伴随着世界各国高等教育进入内涵式发展的历史新阶段，大学评价的理念和方法必须与时俱进。评价体系的科学设计、评价指标的合理选择、评价模型的创新优化和评价效果的价值重估，将成为辅助大学和学科发展的重要方面。为此，需要厘清现行大学评价存在的主要问题，明确学科建设在大学评价中的核心地位，并在此基础上改进和完善学科建设评价体系。

一、大学评价的发展历程

从评价的内容、影响范围和实施规模及水平来看，大学评价的发展历程大致可分为初创期、成长期、成熟期和转型期四个阶段。

（一）初创期：侧重考察学科水平，评价指标比较单一，评价客体规模不大

以大学排名为主要形式的大学评价实践活动最早可追溯到1870~1890年期间美国教育局委员会颁发的大学分类年度统计数据报告。二十世纪美国高等教育的迅速崛起，使得各种大学排名活动变得异常活跃。1925年开展的以声誉为指标对大学学科进行排名的“美国研究生院研究”，被认为是世界上第一个关于研究生质量的排名。1964年由美国教育理事会组织的评估，是美国对研究生院正规评估的开始，也是学科建设评价的开始。国内大学排名起步较晚，1985年《中共中央关于教育体制改革的决定》提出，政府要对高等学校办学水平和教学质量进行评估。1986-1987年组织开展的国家重点学科评选工作是重点学科建设的第一步，也是我国高等教育学科建设评价的一项重要工作。

（二）成长期：关注大学综合实力，评价形式多样，评价影响广泛

一般认为，世界上第一个综合性排名始于1983年美国《美国世界新闻与报道》（简称U.S. News）发布的“美国最好大学”排名，采取对美国领导人调查的方法选出美国最好的大学。自1987年U.S. News连续发布排名之后，1991年加拿大《麦克林》（Maclean's）新闻周刊对46所加拿大大学的文科系和科学系的办学质量进行分类排名，其结果以“A Rating Road Map”为题发表在1992年的《麦克林》新闻周刊上，为高中毕业生择校提供信

息。英国大学排名最早是由《泰晤士报》（The Times）于1992年公布的英国第一份大学指南——“最佳大学指南”（Good Universities Guide），指南包含了101所大学，14项评价指标，内容偏重教学，为中学毕业生申请大学提供信息。德国大学排名起步较晚，1998年由德国高等教育发展中心（简称CHE）与德国商品和服务测评机构（Stiftung Warentest）合作开发第一个大学排名，针对工商管理 and 化学两个学科领域进行评价，主要目的是为学生和高等教育机构提供公正而有价值的大学质量信息。由中国科技信息研究所于1991年发表的中国第一个论文被引次数的大学排名，标志着中国学者对本国大学开展定量排名研究。这个时期，国内官方对大学的评估，有显著社会效应的是“211工程”“985工程”建设高校的认定。

（三）成熟期：注重大学评价研究，评价方法日趋成熟，评价组织更加多元

自2003年开始，随着第一个全球性大学排名——“世界大学学术排名”（简称ARWU）和之后一年《泰晤士报高等教育增刊》（Times Higher Education Supplement）“世界大学排名”的问世，大学排名的实践及相关研究开始进入快速发展期。这时期各国大学排名形式多样，参与的组织机构更加多元，由于信息技术的快速发展，大学排名实施范围更加广泛，世界性的大学排名日益增多。

当前，大学排名受到了世界性的关注，国际性组织、欧盟等都设立了专项课题对大学排名及相关问题开展全球范围内的研究。国际上知名的学术机构、媒体、商业公司，如《美国新闻与世界报道》（U.S. News & World Report, U.S. News）、英国的《泰晤士报》高等教育副刊（Times Higher Education, THE）、英国夸夸雷利·西蒙兹公司（Quacquarelli Symonds, QS）、上海软科以及科睿唯安公司等发布了大学排名评价体系，而在国内大学排名方面影响力较大的有“中国大学评价”（中国管理科学研究院-武书连）、“中国大学排行榜”（香港网大Netbig）、“校友会中国大学排名”（艾瑞深-蔡言厚）、“中国大学及学科专业评价”（武汉大学中国科学评价研究中心-邱均平）等。

(四) 转型期：强化学科质量意识，评价对象更加细化，评价体系不断完善

近年来，以学科作为基本单位的排名逐渐得到了各国政府及教育管理部门的高度重视。例如，澳大利亚研究理事会 (ARC)、英国研究理事会 (RCUK) 下属的英国生物技术和生物科学研究理事会 (SRC)、美国基础科学研究的主要资助机构——美国国家科学基金会 (简称NSF) 均开展了类似的学科建设评价工作；法国成立了国家评估委员会 (CEN)，专门从事学科建设评价；德国虽然没有建立起国家学科评价体系，但其学科建设评价在各州得到了广泛的开展。国内开展的一系列学科建设评价工作，比较有影响力的是教育部学位与研究生教育发展中心牵头，从2002年首次开展，至2017年已完成四轮学科整体水平评估；国务院学位办授权各学科评议组、教指委及省级教育管理部门，从2014年起开展的学位授权点合格评估和专项评估；国务院教育督导委员会办公室委托教育部学位与研究生教育发展中心，2016年首次组织开展的全国专业学位水平评估；教育部等三部委于2017年重点推进的世界一流大学和一流学科建设高校、学科的认定工作。至此，我国由国务院、教育部及其直属部门学位与研究生教育发展中心组织实施，覆盖不同学位授权类型和层次，面向不同学科门类 and 侧重需求的学科综合评价体系已基本形成。

二、大学评价体系差异分析

目前，国内外从事大学评价的机构众多，但认可度较高的排行榜主要有4个，即上海软科世界大学学术排名 (ARWU)、泰晤士高等教育世界大学排名 (THE)、QS世界大学排名、美新世界大学排名 (U. S. News)。在学科评价方面，国内最有影响力的是教育部学位与研究生教育发展中心组织的学科评估。这些排名虽然都能在一定程度上反映大学或其学科的发展水平，但在考察重点和评价思路方面存在明显差异。

(一) 评价类型的差异

不同时期的大学排名形式各异，从排名组织机构看，有报纸、杂志等大众传媒、政府部门和独立研究机构等；从实施的地域范围看，有全球性排名和国家(地区)内排名；从评价内容看，有大学综合实力排名、学科排名和学术水平排名等；从评价指标设置看，可分为单指标评价和多指标综合评价；根据使用的指标类型及其权重赋值大小，可划分成声誉型评价、客观型评价与均衡型评价等。教

育部学位与研究生教育发展中心开展的第四轮学科评估以客观指标为主，声誉指标权重较小，基本上属于客观型评估。

(二) 评价指标的差异

全球型或立足国内的大学、学科排名，主要目的是为公众提供信息、帮助学生择校，为大学参与竞争提供信息和参考。在指标设定上都共同关注大学学科的科研水平，通过论文被引量及高质量论文情况等指标加以表达。但不同评价体系的组织者又有不同的考量，指标的选取存在一定差异，所形成的榜单有一些差别。譬如QS还关注同行评议和雇主评价，ARWU关注获奖校友、教师及高被引科学家，THE则辅以教学水平、产业收入及国际视野等来衡量，USNWR还包含同行评议、学生选择、学生成功以及资源情况等要素。对比发现国外的大学评价更为重视学术声誉和学术指标，国内大学评价体系比较重视教育质量和学术水平。全球性排名主要关注科学产出指标，国内大学排名则更为重视人才培养和资源投入。

(三) 评价方法的差异

评价方法主要涉及不同的价值目的，如何选择评价指标、选择哪些评价指标、指标权重如何分配？欧美等发达国家开展大学评价的理论研究和实证研究较多，可用于文献计量研究的综合性数据库比较齐全，评价理论模型和技术方法比较先进。例如，20世纪70年代以来，《科尔曼报告》引发了对学校效能的探讨，1985年，泰勒(Terry Taylor)等人首次提出增值评价法。1978年，A. Chalmes和W. W. Cooper等人创建的数据包络分析法，具有算法简单，适合多输入多输出复杂系统的相对有效性评价等诸多优点，得到了快速应用。在20世纪80年代初，数据包络分析方法逐渐在欧洲以及北美流行起来。国内的实证研究较少，基本以描述性研究为主。虽然学科评估在评价指标和权重赋值方面也借鉴了国外学科评估的方法体系，但评价数据模型和方法的研究不够深入，至今鲜有完整的理论方法见诸文字。

三、大学评价存在的主要问题

我国的大学评价及排名研究始于20世纪80年代后期。总的来说，它促进了大学之间的竞争、强化了办学水平高低可通过某种标准加以评价的意识，从而提醒高校管理者要重视办学效益，特别是人才培养质量。但客观上看，现行的大学评价方法还有很多不足，如不同类型的院校放在一起比较，

采集的数据来源不一致，所依据的标准和运作的程序不够透明等，就学科评估来说，主要存在以下一些问题。

(一) 评价理念落后时代需求

大学评价归根结底是服务于学科建设工作的，而学科建设是服务于知识的传承和创新，建立以学科为载体的科研平台和组织，培养更多的人才以支撑学科发展，通过学科人才的集聚和价值目标的共同追求，产出更多学术成果，创造更大的社会知识和经济价值。学科建设是人类社会的一种活动形式，真理原则与价值原则是人类活动的两大基本原则，因此学科建设评价要实现学科建设活动真理与价值的双重判断，即建立对学科知识成果真理和实践应用价值的评价标准。学科建设是一个动态发展的复杂进程，其存在和发展既受现实资源条件的制约，又有服务于区域经济社会发展的时代需求；既有学科组织建制体制规模的差异，也有学术环境和文化积淀的影响。对学科建设的评价，倘若不能够准确掌握其内涵和本质，把人为设定评价指标的综合打分作为某个学科建设质量的评定依据，往往是不全面、不合理的，必须树立整体、系统、前瞻的学科建设评价视角和理念，这也是现阶段国内外各种学科建设评价指标体系应重点加强论证的重要方向。

(二) 评价指标不尽合理

质量和效益是反映内涵式发展的主要内容或重要标准，现阶段国际上实施的学科建设评价多采用资源型的目标性评价指标体系，他们往往注重于学科在某一时间段(点)的师资、科研项目、学术成果等的科研水平和学术影响，重视学科建设成效的总量和集成评价，忽视对学科发展潜力和趋势的评价，缺乏对学科建设投入产出的比较评价，未能充分体现学科建设资源的配置效率。从而造成建设基础相对比较差、资源投入相对比较少的学科，学科建设即使取得了很大的进展，在评价结果上却不如基础好但进展不大的学科。一些评价体系的指标设定交叉重叠的现象比较突出，譬如高端师资在人才头衔、学术获奖、团队建设、科研项目等存在一人多绩，同一人的国家级、省部级成果奖项及课题项目均逐项被统计，这样重复的指标数据统计无法说明学科的整体建设水平。还有些指标的选择脱离实际，譬如研究生培养规模受国家或地域招生计划分配导向而导致差异较大，人数的差异会直接影响办学资源和学术成果等评价结果。

(三) 评价主客体关系定位不够清晰

国内外各种学科评价指标体系，普遍重视客观量化的科研成果，譬如学术论文数量、质量和被引，科研成果、项目或获奖等，学术声誉的评价也多采用同行专家主观评价，学科建设最大的利益相关群体——研究生参与评价的内容较少。课程体系结构与质量、导师指导水平与助研金发放、研究生学术活动与科研训练、实验条件及学习资源等内容，这些关涉人才培养的重要方面，研究生应具有最大话语权。因此，优化选择客观科研评价指标的同时，要重视各种参与群体的评价意见，加大学科建设利益相关群体的参与程度。

(四) 评价组织体制机制不够灵活

国内高校普遍缺乏学科建设动态监控的自评预警机制，学科自评和治理能力严重不足，究其原因主要有几点：一是受学科排名的社会效益影响，高校在资源配置上一味偏重于建设基础好、有发展优势的学科，导致资源分配上的简单低效；二是教育行政部门、高校学科数据信息系统建设落后，学科数据监测、评估的技术手段低端；三是高校学科建设体制机制滞后发展，固化的学校、学院和学科(系部)三级管理体系弱化了学科基层组织的学术权利，学科组织的学术功能严重受制于行政权利，基层学科组织丧失了学科建设的主体能动性，导致学科建设的闭环控制系统失灵。

四、完善大学建设评价工作的对策

学科是大学的灵魂，是高等教育的重要载体，学科建设是推动大学发展的关键要素。学科建设状态及指标是体现一个学校在国内外发展水平的重要标志，也是国内外大学排名的主要依据。与大学综合排名相比，学科评价具有更重要的现实指导意义。在“双一流”建设背景下，我国大学发展正面临许多新问题、新挑战，急需有效的大学评价体系予以引领。笔者认为，理念创新是当代大学发展的实践诉求，也是高校学科评价的改进方向，要充分发挥大学评价的导向功能，就必须结合我国高校“双一流”建设实际，从以下6个方面改进和完善学科建设评价工作。

(一) 合理取舍评价体系及评价结论

排名是大学或学科参与市场竞争的一种辅助手段，但不应是大学活动的主要驱动力。众所周知，不论何种形式的学科评价，都需要消耗大量的人力和物力，由于各种媒介的宣传，学科评价产生的社会影响较大，极易引起社会的关注与热议，有

时甚至还会引发争议。如果说大学排名是高等教育竞争的市场化产物，那么学科建设评价活动的产生却是内因和外因共同作用的结果。从高等教育的外部环境看，政府及其部门亟需通过学科建设评价，掌握学科建设的质量和绩效信息，为政策制定提供依据。就高校内部环境而言，受教育资源和政策供给等因素的制约，高校只有积极参与竞争，及时了解学科建设发展的优势和差距，通过富有成效的学科建设举措，针对性地补齐学科建设短板，提升学科建设水平和社会声誉，才能争取到更好生源、师资以及物质资源。从谋求学科长远发展而言，学科建设评价活动能够揭示学科发展过程中的各种矛盾和弊端，有利于促进学科资源优化配置，健全学术组织，完善管理体制，理顺运行机制，从而增强学科发展活力。

公认的学科建设评价侧重点是为大众了解大学学科建设水平和人才培养质量提供信息，帮助学科改善和提高参与竞争的能力。目前，从事学科建设评价理论研究和实践操作的类型主要有三类：

(1)公益学术型。积极关注学科发展态势，以价值建设为导向，一些公益性学术机构或研究人员通过课题研究等形式，研究设计了不同的学科评价体系，促进学科建设评价的理论研究和实践探索；(2)半公益盈利型。通过权威性学科建设评价体系的建立，增加组织机构的知名度和影响力，为社会大众提供学科建设的质量信息，同时也以此为依托为被评价者提供有偿信息咨询服务；(3)管理指导型。高等教育管理部门或资金支持组织为了政策的科学制定和资金的使用效率，建立相应的评价体系，开展有针对性地评价实践，掌握学科建设发展现状和资源利用情况，进一步完善资源配置策略，提升学科建设质量。源于不同社会背景评价主体的不同评价目的，评价指标的选择和综合评价结论会存在较大的差异。比较国内外各种不同的学科建设评价的目标设定、指标选择存在不同的偏向性，评价指标的赋值权重差异较大。为此，学科建设管理者要深入比较研究不同的学科建设评价体系，科学并审慎地选择和借鉴评价结论，避免盲目地迎合评价指标，干扰或影响学科建设工作的正常秩序。

(二)树立正确的评价价值导向

对高校而言，学科建设评价的重要作用是通过多个共性指标数据的比较和分析，认识学科建设存在的问题和不足，促进学科建设主体的自组织能力，从而提升学科建设的质量和效益，实现学科建

设的内涵式发展。为此，学科建设评价指标体系的设计要与学科建设内涵式发展相统一。从历史经验看，学科人才的断代或缺乏将严重阻碍学科的可持续发展。学科建设活动的逻辑顺序首先是利用现有的学科知识和资源开展学科人才的培养。因此，学科建设评价的首要功能就是能够准确诊断学科人才的培养规模、质量和水平，牢固学科的学术本位和人才本位发展意识；其次，学科的健康发展离不开良好的学术氛围，学术氛围的营造需要从学术道德的规范入手。为此，学科建设评价的第二个重要作用是监督和引导学术道德品质的塑造和提升，避免学术失范和学术不端行为的发生，引导学术活动的价值趋同，形成科研成果的水平认同，提升学科建设的内生动力，从而营造风清气正的学科建设软环境；再次，学科建设需要源源不断的资源投入，在学科发展初期，资源的投入与学科建设质量水平基本呈正相关，但随着资源的持续增加投入，倘若学科建设理念、组织管理模式、学科方向设置、人才培养体系不能与时俱进，学科建设绩效的边际效应就会逐渐显现。因此，学科建设评价的另一个重要的作用是引导学科建设主体的积极互动，建立学科评价信息双向反馈机制，通过有效的评价方法和先进的技术手段，客观分析学科的阶段发展水平和建设质量效益大小，促进学科建设和学科评价协同支撑发展。

(三)优化评价的组织管理方式

学科建设评价的模式和方法是动态发展的，是一个不断建设完善和循序渐进的过程，根本上要切合被评价者的理性认识和价值认同，服务于学科建设的实践。国内现有的学科评价主要有学科整体水平的学科评估、学位授权点合格评估、新增学位点专项评估和专业学位水平评估，此外以优先评价为目的的一流学科、重点学科、优势学科的评审立项建设，新增学位点的授权审核，以及学科建设重点项目的年度、中期、验收检查评价等多阶段的学科建设评价。由于学科建设评价的基础性建设不足，每次学科评价都需要收集、整理和审核各种数据信息，增加了学科建设的日常工作压力，学科疲于应付各种评价任务的同时，也因不同评价指标体系的局限性和不同侧重，导致学科建设思路和发展理念不清晰、不稳定，学科内的学术评价标准和科研价值认同不统一。学科建设评价是服务于高校的学科建设，需要培育和增强高校的自我评价意识，建立起学科建设的自我评估机制和管理制度，而各级教

育主管部门的工作重心应该放在高校学科建设自我评价的落实管理和考核上，不宜设置较多名目，频繁干预高校的学科建设工作。

(四)明确政府及其部门的服务定位

众所周知，政府和教育管理部门在学科建设评价活动中起着举足轻重的作用，国内高校学科建设的资源投入主要依赖政府和教育管理部门的支持。为了获得学科建设资源分配结果的多数认同，加强学科建设的管理，保障学科建设目标的实现，必然需要建立一套标准的资源分配体系以缓解竞争产生的矛盾和压力，分配体系的建立必然指向学科评价体系的构建，这也正是开展学科评价的内生动力和外部压力。但是学科是一个集成体，倘若管理部门对学科建设采用单一学术指标和学科排名的片面评价，不仅解决不了学科建设的困境和难题，反而会诱发错误的价值导向，阻碍学科建设的良性发展。斯坦福的经济学家卡罗琳·霍克斯比曾经通过研究发现：政府每给大学增加1%的拨款，这所大学的学术排名就要下降3位，但大学每从市场竞争中多挣到1%的捐款，学术排名反而会上升6位。以ESI学科排名为例，基本科学指标数据库(ESI)是当今世界范围内普遍用以评价学术机构和大学的国际学术水平及影响的重要指标，国内高校争相分析研究和布局进入ESI的学科排名，通过分析发现归属于ESI学科的部分论文是以参与单位或合作单位形式存在的，ESI数据统计的时候并未给予合作者不同的赋值。同时，国内不同机构规模大小、学者研究方向差异、投稿期刊不同偏好，导致同一学科研究论文发表期刊的ESI学科归属差异较大，如此单纯以某一ESI学科的排名来判断学科建设水平，是不够全面的。另外一个值得关注的现象是，一些重大课题、国家级奖励和人才头衔往往集中于学术地位较高的学科或学术带头人，过于看重重大科研成果的获得性评价，就容易造成对学科师资队伍整体水平建设的忽视。学科建设的资源需求是多元的，管理部门对学科建设资源的统筹主要体现在基础资源的配置上，因此政府管理部门对学科建设的评价，应重点在人才培养基础条件的评价和投入产出的效率评价上做文章，其他学科建设资源的投入应通过项目评审立项及课题资助的形式进行调配。

(五)回归第三方评价机构的社会监督本位

学科是一个有机体，学科建设有其客观规律，一个学科的生命力和发展潜力需要多种要素和合而成，既要营造适宜的环境条件，投入足够的经

费资源，还要树立科学的学科发展理念，实施积极有效的管理评价制度。第三方社会组织建立的学科评价，指标的选择主要是建立在对公开数据可准确获取的基础上，没有公开或信息公开不对称的数据，非官方机构难以获得高校的支持而获取，评价结论的可靠性和权威性就难以得到保证。鲜有哪种外部评价能够断定某个学科适不适应社会发展需求，学科组织的凝聚力强不强，学科师资梯队合不合理，人才培养质量的水平达不达标，这些学科建设质量的隐性要素只有通过高校不定期开展的自评估，才能分析掌握和正确判断。政府部门和高校可以适当参考社会第三方评价机构发布的学科建设评价结论，但是不能简单以某个第三方的评价排名来判定学校办学或学科建设水平，否则就会引导高校或学科为了获得有利的排名位次而主动迎合评价指标，导致学科建设生态体系的破坏，还可能会诱发评价机构与高校间的矛盾冲突和利益关联，这就丧失了第三方评价机构的客观中立和社会监督职能。

(六)建立分学科的学科评价体系

学科建设评价指标应重视“知识本身即目的”的原则，切合不同学科的特色属性，聚焦国际标准和学科内涵价值诉求，深入研究和掌握学科内涵建设的一般规律和特殊规律，抓住学科建设的关键问题、并能认真研究和解决问题。要立足学科的集成发展，建立起规范的学科建设评价体系，用以指导学科建设各项工作。传统的学科评价重视的是学科的科研产出情况，未能系统评估科研投入产出率。目前国内学科建设资源整体比较紧张，因此提高资源利用效率尤为关键，要重点加强相对指标的评价，譬如录取率、生师比、师均标志性成果、师均论文引用、篇均论文引用、国际教师比例、国际学生比例、前1%(10%)高被引论文比例、师均出版物、师均科研经费支出、师均技术转让等。

另外，应重点关注学生满意度的调查和评价，依托优势学科建立的良好学术声誉，提高学科整体的社会美誉度，争取更广泛的社会资源用于支撑学科建设发展。

(原文刊载于《研究生教育研究》2019年第12期)

(宫新栋，南京林业大学发展规划与学科建设处副研究员，江苏南京 210037；杨平，南京林业大学发展规划与学科建设处处长、教授，江苏南京 210037)

我国“双一流”建设 实施绩效评价面临的困境与行动路径

杜 瑛

近年来，创建世界一流大学已成为全球性的高等教育竞争性政策之一。实施有效的绩效评价、采取动态机制、增强资源的良性竞争和动态分配正成为国际一流大学建设的主流趋势。当下我国“双一流”建设已转换到有效推进实施的新阶段，绩效评价是审视“双一流”建设成效的重要工具，也是营造公平竞争机制的重要抓手。2020年“双一流”建设的期末评估和调整期限即将到来，国家层面比以往任何时候都更加重视绩效评价和动态调整。尽管以“绩效”为杠杆驱动“双一流”建设目前已经取得各方的共识，但基于绩效的外部评价与大学的自治如何在相互协调中达到共赢，如何兼顾国家战略目标和大学使命、构建科学的绩效评价标准，从而推动不同高校努力实现更高的办学绩效，使得资源投入与国家战略计划的实现相统一，解决好一流大学和一流学科建设支持政策的正当性和合法性问题等，这些都是迫在眉睫需要研究解决的理论和现实问题，也是“双一流”建设政策有效实施的必要保证。国家文件虽然对这些问题有方向性的设想，但具体怎么实施和落实尚无明确说明。目前“双一流”建设实践存在一些误区，未来的绩效评价也将面临一些实施层面的难题，迫切需要从理论到现实的全面审视，以创新性的行这些都动路径和制度设计明确绩效评价有效实施的新方略。

一、绩效评价：新时代提升我国“双一流”建设成效的重要政策工具

随着绩效评价日益成为各国重塑政府公共管理的基本手段，作为公共服务范畴的高等教育领域也逐渐引入了绩效评价的理念和实践。当前，运用公共组织绩效评价的理念管理高等教育已经成为各国

普遍的趋势。我国的“985工程”“211工程”等重点建设的政策假设更多地相信基于行政导向的资源配置的力量，存在准入机制单一、身份固化、竞争缺失、重复交叉等问题，资源使用效益不明显，投资成效受到社会各方的质疑较多。基于行政手段制造出来的重点大学与非重点大学的实力和声誉的差距会诱导非重点大学去追逐和模仿重点大学，从而在某种程度上加剧了大学的同质化。作为“985工程”“211工程”重点建设政策的调适和改革，当下“双一流”建设政策将实施动态调整机制，以绩效评价为手段逐步推进。从2015年的《总体方案》到2017年明确操作策略的《实施办法（暂行）》再到2018年的《关于高等学校加快“双一流”建设的指导意见》都反复强调“坚持以绩效为杠杆”“更加突出绩效导向”“建立健全绩效评价机制”“制定科学合理的绩效评价办法，开展中期和期末评价，加大经费动态支持力度，形成激励约束机制，增强建设实效”等要求，这意味着绩效将是今后国家统筹治理高等教育的一个重要出发点。当下的“双一流”政策更多地转向基于绩效驱动的利益的力量，注重以具有开放性和竞争活力的绩效评价为政策工具，建立激励约束机制，鼓励公平竞争，强调动态调整，突出建设实效。

“双一流”建设中，从人才的投入到办学条件的改善再到办学水平提高与治理能力提升，不仅需要持续大规模的投入，还需要保障资源投入的效率、效果和效益。社会各界对“双一流”建设破解原有重点建设政策中存在的问题、面向未来建设一批世界一流大学和一流学科、进一步推动提升中国高等教育的竞争力等方面都有许多期待。“双一

流”高校在取得必要办学资源的同时如何保持较高的社会公信力？问责其办学绩效就成为政府、社会以及高校自身的必然诉求。绩效评价作为“双一流”建设成效考察和下一轮公共资源分配依据的效率化取向，已成为我国高等教育宏观管理改革与政策调整的重要方向。以绩效评价为政策工具的资源配置机制将成为推动“双一流”建设的新方略。然而，绩效评价作为一种工具性质的政策性号召，面临从政策到实践层面的高等教育运行的现实环境的激励兼容问题，评价的各个环节必然要审视并直面当下“双一流”建设实践中出现的某些偏差和误区，以便切实发挥其价值引领和政策导向的作用。

二、当下我国“双一流”建设实施绩效评价面临的困境

“绩效”导向从“双一流”实施之日起就一直为政策制定者、学者和社会各界所关注。我国“双一流”建设绩效评价的实施尚处于探索期，目前“双一流”建设实践中出现一些误区或偏差，也面临一些实施的难题。新一轮的“双一流”建设绩效评价如果解决不好这些矛盾和冲突，将会影响“双一流”政策的实施成效。

1. “双一流”建设中绩效评价的价值选择面临如何把握大学追求卓越与彰显效益相结合的难题

在“双一流”建设的绩效评价政策中，不同的政策价值，如公平与卓越、成本与效益等之间具有竞争性，虽然都是“双一流”建设政策所追求的，但它们彼此并不能很好兼容，这使得政策制定和执行需要在互为矛盾的价值中做出选择。在推动高等教育改革发展过程中，绩效优先正成为席卷全球的教育管理的主导理念和方向性的行动策略。“双一流”建设的绩效管理背后体现的是“效率优先”的政策价值，政府试图通过绩效评价这一政策工具建立一种动态调整的基于利益的资源配置机制，实施基于绩效的差异化拨款。政府期望财政拨款应该是

竞争的结果呈现，而不是事先的偏向性供给。然而，大学是高投入、高成本、回报周期长的公共服务部门。“高等学校的主要目标是追求卓越、声誉和影响力。……每一所高校都尽力筹集所能筹集的每一分钱，然后把它们全部花光。”大学是一种成本最大化的组织，简单地衡量大学的短期效益与其人才培养的长周期回报之间存在合理性的质疑。基于绩效的“管理主义”的准则与大学的理想、使命及办学目标等也经常会产生矛盾，导致一些评价得不到认同。“双一流”建设中如何把握大学追求卓越和彰显效益的统一，成为绩效评价推进实施中面临的难题之一。

政府将“双一流”建设作为重大战略决策，其目的是提升我国高等教育的综合实力和国际竞争力，建设高等教育强国。政府的期望实质上隐含了对绩效评价的价值选择，体现了对绩效评价本质的理解和把握。政府始终出于一种管理的需要，期望用绩效评价作为管理的手段和工具，绩效评价结果要成为改进管理的重要依据，而较少是出于高校自身改进的考量。大学自治和学术自由是大学的理想与精神，是高等学校努力追求的自身作为主体的需要和价值，问责则意味着某种程度的外部干预。政府指标化的绩效考核导向可能对大学学术研究活动形成倒逼效应，使得大学学术研究活动的内在规律被行政管理、政府政策的功利性所替代，从而对大学的自治带来严峻挑战。在实施绩效评价时，如何正确利用绩效杠杆，使基于绩效的外部问责与充分尊重大学的理想及自治诉求之间保持适当的平衡，避免过分追求量化的绩效指标，这需要进行政府、大学和市场共同参与的高校内部和外部治理制度的综合设计。

2. “双一流”建设评价准则的确定面临如何弱化基于排行的唯科研指标倾向与彰显人才培养过程增值的现实需求的难题

“双一流”建设存在基于排行的“唯指标

化”倾向。“大学排行和类似的一些分类工具作为一种公共信息机制，正日益成为衡量和比较第三级教育机构绩效所依赖的方式。”目前我国尚未建立科学的“双一流”评价体系，这使得一些地方政府和高校不得不跟着并无充分科学依据的大学排行榜走。调研发现：一些地方政府和高校正在将世界大学排名和学科排名作为追求的关键指标，将提升排名作为其战略目标之一。很多高校领导在汇报一流大学或地方高水平大学建设目标或成效中，直接将学校的ESI学科排名、SCI文章等产出指标作为衡量办学绩效的重要指标，而鲜少提出人才培养方面的目标及具体发展情况。世界排名正在成为公共资源及“双一流”政策效果的担保者，不排除相关主管部门在对高校进行绩效评价的时候，仍然会将一些大学排名、学科排名作为至关重要的依据，进而使得相关指标成为引导大学发展的“指挥棒”。排行榜只是对大学发展中可量化结果的某一种测量标准，不同的排行榜是基于对大学意义价值的某种理解，基于这样的理解选择评价指标本身就很有争议。在排名体系下，对于一流的认知更多地取决于市场和商业的逻辑，容易使人们将建设世界一流大学视为对办学资源的争夺和办学绩效指标的攀比，而忽视对涉及大学精神和大学使命等根本性长远问题的关注。目前我国的“双一流”建设也受此影响。大学可能为了提升其排名而将重心放在科研产出上，即使教学是其主要使命。对关键科研产出指标表现的追求由工具上升为目的。“很多大学将大学发展的重心转向了排名竞争，资源配置向有利于提升大学排名的指标倾斜，”一些高校甚至开始盲目地遵循排名指标来进行学科建设，将全部资源倾斜于少数应用学科和生物、医学学科等这些比其它人文学科更能从排名中获益的学科。部分学校对相对弱势学科如教育学等人文学科进行了不同程度的调整或裁撤，一些人文学科正在陷入危机中。

当大学专注于“关键绩效指标”时，它将注意

力集中在衡量产出，而不是包括培养和关怀在内的教育过程和投入……，与此相关，在高校内部治理中，提升科研绩效也通常被认为是提升声望或声誉的有效手段。基于科研绩效的评价引导机制使教师在晋升晋级时受制于主持纵向项目的级别、获取研究经费的多少、发表论文的数量及刊物的级别等“硬性要求”。这样的制度环境和学术生态诱发了一些功利主义倾向和学术异化现象。“教师们迫于压力，需要在短期内发表大量学术成果，往往根据学术期刊和出版社的喜好而写作”。人才培养是大学赖以存在的基础，也是一个长期的过程，很难用看得见的量化指标来衡量，如果我们在绩效评价实施过程中唯科研指标化，忽视人才培养、大学精神和现代大学制度等方面的培育，可能会影响一流大学长期战略目标的实现。如何彰显人才培养过程增值的现实需要、弱化基于排行的唯科研指标倾向是“双一流”建设确立绩效评价准则时迫切需要关注的难题。

3. 绩效标准的确定面临如何避免不同类型高校价值诉求的多样性与绩效指标和结论统一性的矛盾与冲突

尽管以“绩效”为杠杆驱动“双一流”建设目前已经取得各方的共识，但什么是“绩效”是有争议的，如何统一衡量“绩效”也一直是一个难题。大学也许是当代社会最复杂的机构，“对使命不同、资源状况不同的院校进行比较本身就有方法上的缺陷和不当之处。用一种价值标准衡量使命、任务不同的众多高校更是一件极不合理的事情。尽管不同高校和学科的“投入”“过程”和“产出”的总类型是一样的，但具体的投入、过程和产出因素的比例结构有很大不同。不同类型高校和学科确立共同的绩效标准显然是行不通的。如何界定“双一流”建设中不同高校及学科的绩效，如何取得对绩效评价的共识，将各方认同的评价结论用于资源配置，从而推动不同高校努力实现更高的办学绩

效，是“双一流”建设实施绩效评价迫切需要明晰和解决的问题。

国家层面提出“双一流”建设目标后，几乎每个省（市）均提出了地域性的地方高水平大学建设项目，“双一流”建设已波及整个高等教育系统。调研发现：目前多数省份提出的“双一流”建设目标是按照建设世界一流大学的思路来建设地方高水平大学，没有很好地体现地方“双一流”建设的目标、任务和内涵发展要求，而且可能会带来高校发展愈加同质化，威胁整个高等教育系统多样化的生态。地方高校盲目地跟着进入国家“双一流”的大学及学科在发展目标中追求ESI学科数量和SCI文章，弱化了地方高校更应该根据当地经济社会发展、产业需求和市场需要及时优化学科专业发展方向的办学要求。如何综合考虑国家战略、地方区域经济社会发展需求与大学使命，充分考量不同高校之间的差异，构建既能反映每所高校的进步程度又能根据评价结果进行拨款的绩效标准，引导不同高校的特色彰显和使命实现，是“双一流”建设中绩效评价面临的难题。

4. 评价主体的确定面临管办评分离导向下单一行政性评价主体为主与多元利益主体博弈的矛盾与冲突

《总体方案》提出：“建立健全绩效评价机制，积极采用第三方评价，……”《实施办法》提出：“参考有影响力的第三方评价，对建设成效进行评价……”；《关于高等学校加快“双一流”建设的指导意见》指出：“鼓励第三方独立开展建设过程及建设成效的监测评价。”这些规定表明，社会评价机制成为政府用来引导、监管高等教育发展的工具，教育行政部门希望更多地采用第三方的评估结果，以促进政府管理方式的改革创新，而不是过去那样基于行政手段的直接下达行政命令或指令的评价。但目前我国第三方评价的主体尚未得到非常清晰的界定，隶属政府的评估机构存在主体身份

的复杂性问题。这些评估机构依靠与行政部门的隶属关系进行“职能转移”或“直接委托”来获取教育评价项目，同时享有从政府部门获取各种高校办学信息的便利性；评估活动并不是价值无涉地运作，而是体现了政府的政策意图。其评价程序、评价标准和关键性的指标也主要征求教育行政部门的意见，是建立在政府政策偏好上的运作架构，最后的排名与淘汰的原则很大程度上也是用来帮助政府实现其意图的工具，是政府的职能延伸或政府的代理人。当下对“双一流”建设资源分配影响较大的国内评估机构多依附于政府或是政府所属的事业单位。如：从国家层面来看，教育部学位中心开展的学科评估是国内大学比较看重的学科评估，由于它是教育部直属的事业单位，具有获取各种大学办学信息的便捷条件，其评估结果被政府用来参考作为对高校进行资源配置的主要依据，如2016年下拨的“双一流”建设经费分配依据主要是第三轮学科评估的结果；从区域探索来看，上海对高峰高原学科的阶段性评估是上海市教委委托其直属事业单位上海市教育评估院做的，上海市教委公布的《上海高校高峰高原学科建设第二阶段动态调整学科名单的通知》对高峰高原学科实施了动态调整，直接依据就是上海市教育评估院2017年启动实施的上海高校高峰高原学科建设第一阶段绩效评价结果。

目前我国专业化、科学化、规范化的社会中介评估机构还远未成熟，尽管民间的评估机构纷纷开展评价活动，但评价范围和社会影响有限，其方法论基础颇受争议。市场主导的商业性大学排名机构在独立性、科学性和可靠性方面还面临着质疑。整个高等教育行业协会和专业组织缺乏合法渠道表达他们对“双一流”建设成效的关心，社会及用人部门缺乏参与评价的途径和动力。“双一流”建设的绩效评价与高校获取的办学资源密切相关；同时，绩效评价涉及多方利益，管办评分离背景下如何兼顾单一行政性评价主体与多元利益主体的利益是绩

绩效评价政策实施面临的一个难题。如何参照国际标准和通行评价办法培育或引进独立的第三方中介机构、建立健全多元利益主体参与的绩效评价机制是迫切需要解决的问题。

三、以绩效评价为政策工具破解“双一流”建设困境的行动路径

未来绩效评价的实施需要超越传统的路径依赖，进行从绩效观、评价准则、绩效标准到评价主体的全方位变革，实现各个层面的激励兼容，以创新性的行动路径和制度设计保障和激发作为办学主体的高校努力实现更高的办学绩效。

1. 确立关注人才培养过程的绩效观，充分发挥绩效评价动态监测的诊断和激励功能

“双一流”建设中的绩效评价是以绩效的理解为基础的，绩效内涵的广泛认同是形成评价标准的基础，其核心要素直接影响评价体系的构建。现行“双一流”评价中所出现的问题正显示了一些地方政府和高校在绩效理念上的认识不足，以致造成建设中的盲目竞争现象。“双一流”建设迫切需要在观念上确立关注人才培养过程的绩效观，在实施环节上树立杜绝短期功利主义的政策取向。新时代“双一流”建设的绩效导向应包括三方面：第一，基于输入绩效的评价，体现基于条件和目标的资源竞争；第二，基于过程的绩效监测，主要目的在于警示、督促和动态监测“双一流”建设的过程，关注人才培养过程的增值功能，体现对高校发展和学科建设的诊断及激励功能；第三，基于结果绩效的评价，具体包括两方面：一是评价结果向社会充分公布，以公开信息的方法来影响高校的社会声望，评价结果作为质量监控的反馈机制，要为高校和学科的质量改进提供信息支持和保障，二是依据基于结果的绩效证据，建立投资的追回和奖励机制，激励和问责高校的建设责任。

基于以上认识，“双一流”建设中的绩效评价是通过对高校或学科建设的投入、过程和产出的衡

量，对一定时期内的建设效益和业绩作出综合的价值判断，以检验建设成效和决定后续资源配置的流向。“投入”主要是各种办学资源的投入，包括公共财政、人力资源和其他公共财产的投入，有时也把政策许可等作为投入要素来看待。对“过程”的监测评价可考察学生的学习体验、学生的成长与发展情况、教学质量等方面。对于大学“产出”的衡量，除人才培养、科研创新和社会服务外，对社会精神的引领、对文明发展的促进等这些大学的外部效应也是重要的衡量指标。

“建设成效的实际效果取决于一个国家对于成效的理解是如何在政策议程中得以体现并立足。”

“双一流”建设的绩效评价本质上既是一种问责与监督机制，又是一种激励机制。绩效评价作为监测“双一流”建设成效的工具，有对学校和学科进行动态监测的功能，要更多地发挥其发展性和诊断性的功能。人才培养虽然作为“双一流”建设的重要主题之一，但往往因周期长、见效慢和难以被具体量化而在实施方案中被边缘化；“双一流”建设中的绩效评价要以当下的“破五唯”的政策导向为契机，建立关注人才培养过程的绩效观，将过程性指标和结果性指标相结合，适应新时代对教育价值导向及评价导向的新要求，“更多采取贡献和过程的绩效观，注重高校自身在实现其责任时的主体性，如此，方可真正实现办学自主与绩效问责并行不悖的双赢局面。”

2. 建立协商沟通和责任共享机制，多元主体协商确定绩效评价标准与指标

绩效指标设定是一个达成共识的过程，要充分尊重学校自主权。虽然建设世界一流大学是政府和大学的共同愿景，但两者关于“世界一流”的认识和理解并非完全一致，政府大量财政资金投入背后更多地是基于高等教育强国的考量，虽然政府也强调尊重大学的学术自主权，但政府的行为受自身利益诉求和官僚逻辑的驱动，在提供巨额财政投入的

情况下不会轻易放弃在“双一流”建设中占据主导地位地位的诉求。地方政府建设高水平大学本身也承载着促进区域经济发展、产业升级和繁荣文化等方面的利益诉求。而大学更多地关注办学使命的实现、学科的发展和学术的传承与创新。鉴于此，建立多元利益主体的协商沟通机制和责任共享机制，有效调节政府、高校和社会之间的利益诉求和价值主张势在必行。“双一流”建设的绩效评价制度变革的走向应是建立绩效标准及指标的沟通协商机制，让多元利益主体参与到评价和问责的议程中。绩效标准的确定应在政府、高校和教师等多方利益主体协商、沟通的基础上生成，形成基于价值认同的绩效评价标准。最终使学校或学科建设成效的实质性改进成为各个层面的关注对象，使提高“双一流”建设成效成为政府、高校和社会的一种共享责任，构建政府、大学和社会各有侧重、互相衔接的共治型绩效管理引导新机制，引导高校对照建设目标努力提升办学效果及效益，让绩效评价作为对这种结果的检验。同时，高校要主动提升责任感，建立内部监测评价制度，及时发布年度建设进展报告，落实其主体责任。

“大学的本质是高等教育绩效评价的灵魂。”具体的绩效评价标准的选择不应局限于量化指标，还要考量制度、理念和大学精神等支持大学真正走向世界一流的要素，在大学使命、大学精神、制度建设和人才培养等方面要有明确的引导性体现。对于这些不易量化评估的方面可采用定性的同行评议的机制进行，以概括性问题的方式确立明确的评价标准。

3. 建立分类的绩效评价制度，将“双一流”建设的教育治理导向大学发展的多样化

“双一流”建设的绩效评价要适应高校多元化的现实，建立分类评价制度。评价准则的确定要与不同类型大学的使命和人才培养目标紧密结合，即评价准则要综合考虑国家战略需求、区域经济社会

发展需求和高校自身的发展使命，绩效指标的选择既要考虑每一类高校的共性，又要体现对不同高校个性和特色的尊重，将“双一流”政策中的教育治理导向大学发展的多样化。不仅对高校和学科的评价要坚持分类的原则，大学内部绩效监测评价也要结合不同类型高校的使命与职能构建分类、分学科的符合不同岗位教师劳动特点的评价制度。

不同类型高校承担的使命与职责、愿景目标、人才培养类型、核心学科和规模体量等都有自己的特点，学生的学习成果表现也是不同的，无法用统一的标准来衡量投入和产出。科学的绩效评价标准的制定必须立足于对不同类型大学人才培养类型和核心使命的理解上。罗伯特·鲍尔在考察若干院校实践的基础上，建议要改变排行榜的方式，院校应采用符合其使命的绩效指标以更好地促进自身发展。具体的绩效指标的选择要综合考虑不同类型高校的人才培养主体功能的目标定位、大学传承与生产知识的类型、承担科学研究的类别、社会贡献及服务面向等维度。如大部分国家部委直属的“双一流”建设高校是以知识创造和培养拔尖创新型人才为取向的研究型大学，这类高校的绩效标准应以创新型人才培养为基础，把基础研究与教学的结合与成效、科技创新与知识创新的成效、围绕国家战略和社会服务前沿的重大原创性成果、在文化与社会发展重大命题上作出的杰出贡献等作为重要方面来考虑，引导此类高校引领经济转型、产业革命与文化创新发展的潮流。地方“双一流”高校或纳入地方高水平大学建设范围的高校的评价准则要体现对地方经济社会发展服务的引导，评价标准要关注知识的应用，强调教学与科研并重，将产教融合和区域发展的密切度作为衡量办学绩效的重要依据。地方高校的学科专业宜遵循社会服务导向和社会需求逻辑，而不是纯粹的学术导向，要强调对区域经济社会发展的贡献度，引导地方高校依托优势学科，主动融入区域经济、产业和社会发展，使其人才培

养与产业、人力资源需求形成良性互动。绩效评价的标准要释放出“支持不同高校在各自领域和类型中争创一流”的政策导向，从而提升高等教育的整体发展水平。

4. 建立政府、高校和第三方专业机构等多元评价主体的协同参与机制，综合多方评价信息反馈推动高校不断取得更高的绩效

行政体系过多介入不会造就世界一流大学或学科。除资金外，治理是高等教育系统绩效的另一个关键决定因素。从评价主体来看，随着管理型政府向服务型政府的转变，政府势必要从评价中抽身出来，建立政府、高校和第三方专业机构等多元利益主体协同参与的共治型评价机制势在必行。国际上一流大学建设政策得以初见成效的必要保障是成立专门的审议委员会，负责建设计划的评审、对计划实施的监督和总体绩效评价，以此平衡政府与高校间的利益关系和治理格局，保障财政拨款和评价的客观性与公正性。如德国的精英大学计划是基于学术自治原则的政府行为，其能够实施动态建设的关键在于由德国科学基金会和科学审议会联合组建共同委员会，成立了权威和国际化的审批委员会，构建了一套完整和公正的评审机制，项目的运行过程并不受政府控制。我国台湾地区的“发展国际一流大学及顶尖研究中心计划”和“迈向顶尖大学计划”的审核和评估均由专门的审议委员会执行，除了对项目申请阶段进行严格评审以外，还负责对计划实施的总体绩效进行整体评价，以评估计划实施的效果。香港建设世界一流大学的政策制定、审核、实施、经费拨付等均是教资会等中介机构负责，教资会在维护高校的自主和平衡政府与高校间的关系上起到了至关重要的作用。我国“双一流”建设可借鉴国际经验，加强整体性制度设计，在国家层面设立专门的审议委员会，建立监测评价的数据信息平台。政府作为高等教育的宏观管理者，主要致力于为“双一流”建设提供良好的信息

服务和有序的竞争秩序等。审议委员会成员构成要包括政府部门、高校或科研机构、行业组织人员及国外相关领域人士的各方专家代表。具体评价活动的实施可以委托独立的第三方专业机构进行，发挥专业机构在院校评价和学科评价中的专业导向作用。但我国目前尚缺乏独立的第三方评估机构，可探索培育具有独立法人资格的第三方评估机构，如采取改造准官方的教育评估机构、大力发展学术性评估组织、逐步培育高校等自治性的评估机构和积极扶持民间非营利评估组织等不同方式多管齐下培育独立的第三方评估机构。

从评价信息来源和使用来看，在对“双一流”建设进行绩效评价时，应综合多方评价信息和结果，绩效评价信息收集除了教育行政部门自身掌握的相关信息外，还要充分采用院校自身、院校的利益相关者和独立的第三方评价机构等多方来源的支撑性信息。具体来讲，可考量以下几方面的信息来源：一是院校自评，即进入“双一流”建设的高校或学科作为被投入对象所提供的能够证明建设成效和绩效的证据信息；二是多元利益主体，包括教师、学生、校友、校企合作办学者等作为评价主体参与的对相关院校或学科的评价信息；三是行业协会、专业学会、第三方评价机构独立评价后所提供的科学客观的绩效证据。将以上几方面的综合评价结果与国家绩效监测平台的相关信息综合起来共同作为对“双一流”建设绩效进行价值判断的信息来源。最后根据绩效评价结果对“双一流”高校和学科实施周期性、动态化的资助投入，形成滚动淘汰机制和良性循环机制，从而建立健全多元主体参与的绩效评价的长效机制。

（原文刊载于《教育发展研究》2020年第3期）

（杜 瑛，上海市教育科学研究院智力所副研究员，上海 200032）

世界一流大学五大评价指标体系的比较、改进及其启示

赵国栋 马瑞敏

一、发展背景与学术研究动态

2017年10月18日，习近平同志在党的十九大报告中强调，“建设教育强国是中华民族伟大复兴的基础工程，必须把教育事业放在优先位置”，要“加快一流大学和一流学科建设，实现高等教育内涵式发展”。对于大学的建设与发展而言，科学客观的评价标准具有两方面的显性功能：一是通过科学的评价方式找到一流大学办学水平高的客观因素；二是通过评价找到大学之间的差距，在实现良性竞争的基础上弥补不足，促进大学的发展与提高。显然，梳理一流大学的发展背景和相关评价研究的基本情况，能够让我们对于建设世界一流大学的真正内涵和着力方向形成基本共识。

（一）世界一流大学的发展背景

自现代意义上的大学诞生之后，每个时代都有备受世人关注的大学，无论是西方的博洛尼亚大学、柏林大学、约翰·霍普金斯大学、斯坦福大学；还是我国近代的京师大学堂、北京大学、西南联大等，都处在当时的塔尖地位，成为其他大学和后来者不断模仿和研究的对象。近十几年来，建设世界一流大学的呼声在很多国家持续升温，世界一流大学成为惯用词汇。然而，西蒙·马金森认为，那些名列前茅的大学极少使用该词汇，该词汇在新兴国家更受关注。按照这样的理解，我们认为世界一流大学是一个历时性概念，也是一个共时性概念，应当是某一时期在某些方面表现更为优秀的大学。

在21世纪的语境中，世界一流大学就是指以美国研究型大学为样板的大学，也就是说，世界一流大学应当是以学术研究为核心并表现杰出的大学。研究者们普遍认为，世界一流大学建设炙手可热的缘由在于：无论从教育发展还是人类历史的视角，国家的经济增长从来没有像今天这样依赖于教育的作用。就教育的经济和社会价值而言，20世纪的舒尔茨、罗默、卢卡斯和21世纪以来的希克斯、巴罗等学者的研究对此已经进行了充分的论证。事实上，除大学一以贯之的超越性发展内核外，经济全球化、高等教育国际化和高等教育在国家软实力竞争中的价值正是其上升为国家战略的三个重要原因。

随着大学对科学技术发展、人文精神引领等与社会进步、国家战略相关事业贡献率的逐渐提升，建设卓越的大学已不仅是关乎大学自身的发展问题，而是许多国家不约而同的共同的主动选择与战略决策。对于中国而言，除了高等教育在国计民生中的重要性日渐增强，并以此受到国家和政府的积极推动外，也有新时代给予了大学难得的发展空间等因素，大学因此有机会进行发展逻辑的理性反思，更有寻找建设大学路径并使其不断卓越的雄心壮志。

（二）世界一流大学评价研究动态

为了考察学术界对世界一流大学评价研究的基本动态，笔者通过中国知网CNKI这一检索工具，以世界一流大学评价为主题，检索得到相关文献246篇；以世界一流大学评价为篇名，检索得到相关文献53篇，文献的发表年度基本上都在2000年以后。基本的研究范围主要集中在世界一流大学内涵、科研竞争力评价、基本科学指标、评价研究等领域。可以看出，国内学者对于世界一流大学的评价进行了卓有成效的研究。

在已有研究中，可以分析出四点共识：其一，指标体系方式，实质上是把大学的所有工作归结为客观的、单一的可比性统计，在某些学科领域意义显著，而在如人文社科类的学术领域意义甚微，甚至可能出现负效应；其二，大学排名研究是舶来品，我国的大学排名指标体系属于“洋为中用”，我们从“双一流”建设名单及其相关信息，看到了诸多国外大学评价体系的核心内容，或者说目前的评价标准以国外一流大学的评价标准为基础；其三，国内的研究更多将焦点放在了指标的合理性上，强调本土意义上的学术可比性求同，关注指标与质量之间的逻辑关系；其四，大学排名的社会影响外在因素大于大学内在的纯粹意义。

结合教育部倡导的大学动态调整策略，我们认为，具有中国特色的高校“双一流”建设和评价标准仍需进行持续不断的探索。正如清华大学史静寰教授指出：中国顶尖大学已经逐步从形似（科研产出等硬性指标）向神备（中国特色）的方向迈进。这其中包含着无数教育者和研究者的美好愿景：到21世纪中叶，中国有若干所真正达到

神形兼备的一流大学，无论是基于国际的软硬评价指标，还是在人们的感性理解中，我们都能做到一流。从目前世界大学发展趋势而言，欧美的顶尖大学是无可否认的世界一流大学的象征和存在，所以，我们的大学评价在很大程度上都是在寻找和学习它们的共性，求同与同质化超越仍是新时代中国大学发展的长期目标与主题。

正是基于以上背景，我们选取了国内外认可度较高的五种世界性大学评价体系作为研究对象，分别是：上海交通大学世界一流大学研究中心进行的世界大学学术排名” (Academic Ranking of World Universities, 简称ARWU)、Quacquarelli Symonds世界大学排行榜(简称QS)、中国科学评价研究中心的Research Center for Chinese Science Evaluation(简称RCCSE)、《泰晤士报高教副刊》(简称THE)、《美国新闻与世界报道》(U. S. News&World Report, 简称USNWR)。本研究运用比较分析法指出各评价体系数据分析下的特点和差异，并对现有评价标准的优势与短板以及以后的发展方向提供建议。

二、五大评价体系的基本情况及其目的

大学排名最早可追溯到1870年，截至目前，在世界范围内已经有超过45种较大影响的大学评价体系。今天学界广泛关注和讨论的大学评价则开始于1983年的USNWR，此后大学评价逐渐成为大学提升办学质量的全球化风向标，形成了当代独特的以技术理性方法为主导思想的高等教育评价范式，至今已有35年历史。在此期间，各种评价机构在世界范围内日渐增多，不同评价标准与评价理念的指标体系经过创新与完善的积累，已经成为大学办学质量鉴定和社会反馈的重要参考依据(表1)。

(一)ARWU

2003年，上海交通大学世界一流大学研究中心和高等教育研究所的研究人员，首次独立发布

数据来源主要为美国期刊和论文平台，只关注国际认同且可比的科研成果和学术声誉，基本上采用可以完全量化、重复验证的客观数据作为主要数据，随后的几次改革，开始探索不同学科的可比性排名，但是仍然以可以量化的学术成果为唯一的评价方式。ARWU建立的最初目的是为了数据化呈现中国大学与世界一流大学间的距离(尤其是科研差距)，客观呈现中国大学在国际大学体系中的位次，目前主要作为政府、大学、国际间留学人员了解中国大学办学水平的窗口，以及政府和高校制定政策的重要参考，引导中国大学向世界一流大学看齐。

(二)QS

QS先后与不同的机构合作，目前形成了包括亚洲大学排名、拉美大学排名等四大跨国型大学排名体系，也是唯一获得联合国教科文组织(IREG)国际专家认证的排行榜。QS因涵盖范围极少和极其注重声誉评价而饱受争议，其评价指标比较简单，雇主评价和同行评价指标权重达到50%。在历年来的指标调整中，QS始终坚持主观评价的重要性，赋予了较高的指标权重。QS更重视大学的社会服务功能，瞄准了全球性高等教育服务市场，以及国家级的国际性贸易服务对于高等教育的需求和认可，比如：随着互联网的兴起，将“学术引用”“网络影响力”作为重要的指标依据，力求其评价范围更为接近雇主的需要和社会的朴素期望。近5年来，中国部分高校因为国际化程度和人才培养质量的不断提高，声誉指标进步相对显著。

(三)RCCSE

RCCSE创立于2002年，武汉大学中国科学评价研究中心在吸取其他评价体系成功经验的基础上，2006年开始进行以科研竞争力为主要内容的世界大学评价。2016年，在高校“双一流”建设背景下，RCCSE打破过去的科研评价局限，对评价体系进行了方向性调整，着重对大学的综合实

表1 五大评价指标体系的基本概况

机构名称	成立时间	评价目的	评价对象
ARWU	2003	中国大学与世界一流大学的科研差距	全世界大学
RCCSE	2002	促进中国大学有效管理决策、竞争发展定位与管理创新保障	全世界大学
THE	2004	为学生选择学校提供信息、为大学的国际化竞争提供参考	全世界大学
QS	2009	为学生选择学校提供信息和就业指导	全世界大学
USNWR	1983	为学生选择学校和政府提供信息	全世界大学

一流大学评价信息，也是全球首先公开发布世界大学排名的机构，主要针对综合性大学尤其是理科和工科大学的科研成果与学术能力进行评价。

力进行评价，此次改进主要体现在通过分类分层的方式对大学进行综合性评价，通过分析指标中既有存量和增量间的进步，以此体现大学的发展

状态。目前，该评价体系主要作为我国政府分析科技创新与进步形势的决策定量支持依据，也是我国大学管理部门进行大学竞争力分析的参考依据，当然也是学生国际留学的信息依据和其他国家与地区大学发展的数据依据。

(四) THE

THE评价机构的前身可追溯到1992年的《泰晤士报》(The Times)。2004年，THE与QS合作第一次发布了大学排行榜，2009年后双方因理念不同正式解散，并各自独立进行大学评价。此后THE对评价方法和评价标准进行了较大的调整，通过与汤森路透科技信息集团合作获取相关信息数据，新的评价指标体系更为复杂，但仍保留了同行评价指标，增加了部分新指标，降低了主观评价指标权重。2015年之前采用web of science作为数据来源，2016年后开始使用爱思唯尔

了这个功能的基础上，又逐步成为政府、雇主、大学和社会制定政策、了解大学发展状况的重要参考。

三、五大评价指标体系的比较分析

(一) 分析方法

本研究所选取的研究内容，如指标体系构成、权重分配等均来自各评价机构的官方网站或者已发布的相关报告。如表2所示，不同的评价机构虽然使用了不同的标准进行排名，但是主要涉及了教学水平、师资质量、科研能力、知识转化收入、国际化水平和声誉等6个维度。其中，教学水平被4个评价机构纳入一级指标体系，科研能力被5个评价机构纳入一级指标体系。从各大评价体系的指标发展和完善历史看，体现了大学质量观的发展和进步，评价的目的从以社会因素为主向大学的内涵式发展转向趋势明显。五大

表2 五大评价体系的一级指标

ARWU	RCCSE	THE	QS	USNWR
教学水平	教学水平	教学水平	教学水平	科研能力
师资质量	师资质量	科研能力	科研能力	科研影响力
研究产出	科研能力	论文引用	国际化水平	声誉
人均学术绩效	声誉	国际化水平	声誉	
		知识转化收入		

Scopus数据库。在2016年发布的评价指标中，THE降低了同行评价的权重，科学研究的权重达到60%，增加产业收入为一级指标，突出大学知识转化的能力，更直观体现出大学产学研结合发展的态势，以及大学对现代生产、经济和科技的正面影响能力。THE的主要目的是帮助各个国家、地区的学生甄别出不同大学之间的差异，进而从中选择更适合自己的大学。

(五) USNWR

USNWR是公认的首先开创现代大学评价的组织，从1983年开始通过邀请美国国内的校长同行评议的方法，对美国本土大学和机构实施排名，全部采用主观评价的方式。2009年开始与QS合作，运用卡内基基于促进会构建的高校分类办法对大学进行分层分类，针对同类型大学进行综合比较，使用了更为科学的调查分析办法，2014年扩展为全球500强大学排名。在历次的指标修正中，从最初的完全定性评价——声誉调查，逐步过渡到定量指标权重占75%，其中对于教学质量——毕业生、优秀生源保有率的表现更为关注，之后增加了校友表现、生源质量和市场对毕业生的认可程度。USNWR最早的服务对象是给学生和家长提供选择大学的信息依据，此后在保留

评价体系的指标虽然各不相同，从不同方面体现了评价的权威性，但是，我们仍然可以从中找到不少有意义的比较内容。

根据五大评价体系各自发布的二级指标及其权重分配，为了能够客观地表现出世界一流大学所关注的主要评价内容，消解评价机构的价值偏好，以及受一级指标对应不同二级指标的困扰，本研究将各大评价体系的二级指标按照教学水平、师资质量、科研能力、知识转化收入、国际化水平和声誉等指标数和权重予以统计，同时，引入了全球可比指标个数和权重作为参照，得出各类型指标个数及权重分布(表3)，以便更直观地比较各大评价体系间的不同特点。

全球可比指标是指能够反映大学在世界大学中的竞争力的指标。比如：诺奖或菲尔兹奖人数、进入ESI排名学科数、被SCI和SSCI收录的论文数和高被引论文数等是评价一所大学教学水平或科研能力的标准；国际教师和国际学生比例等国际化程度指标则被当作全球竞争性的标准；而研究经费、专职教师数量、师生比等指标，在不同的国家具有不同的统计学意义，作为全球性可比指标在一定程度上是片面的；出版物、学术会议和书籍因为语言、版权等问题，各国评价指标

并不相同，也缺乏全球性可比意义。

此外，各类声誉调查虽然被各大评价机构所重视，但是因为其信息和数据来源不同，比如THE和QS的声誉调查，基本上是以发达国家的雇主评价为主，并不能体现国家间的平衡，相比较而言，RCCSE注重网络影响力的指标方式较为客观，QS的学术同行评价实质上体现了大学作为学术共同体间的互相认同，具有客观的借鉴意义。

(二) 分析结果与讨论

通过表4的比较分析可以看出，五大评价体系所认同的世界一流大学的特点主要包括以下五

科研特征，只对某些特定的研究范围和研究成果具有客观的评价意义。

第二，国内的评价体系还没有形成具有中国特色的评价认识。通过消解一级指标的价值偏好，以及结合ARWU和RCCSE历年来对评价指标的调整历史发现，政策的长效影响对于评价指标的干预较大，评价标准力图呈现评价体系的主要关注点，个别维度偏好显著，如语言偏好、刊物偏好。在师资质量指标上，ARWU和RCCSE把教师获得的荣誉称号作为评价标准，而其他三个评价体系则更为关注教学水平。事实上，师资质量应该

表3 五大评价体系中各类型指标个数(个)及权重(%)分布

	ARWU				RCCSE				THE				QS				USNWR				
	①	②	③	④	①	②	③	④	①	②	③	④	①	②	③	④	①	②	③	④	
教学水平指标	0	0	0	0	2	17	2	17	4	15	0	0	1	20	0	0	0	0	0	0	0
科研能力指标	3	60	2	40	4	34	4	34	3	42	2	36	1	20	1	20	9	65	6	50	
师资质量指标	3	40	3	40	1	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
知识转化收入指标	0	0	0	0	1	6	1	6	1	2.5	1	2.5	0	0	0	0	0	0	0	0	0
国际化水平指标	0	0	0	0	1	21	1	21	3	7.5	3	7.5	2	10	2	10	1	10	1	10	
声誉指标	0	0	0	0	1	10	1	10	2	33	0	0	2	50	1	40	2	25	1	12.5	
总计	6	100	5	80	10	100	9	88	13	100	6	46	6	100	4	70	12	100	8	72.5	

注:①=指标个数;②=指标权重;③=全球可比指标;④全球可比指标权重

个方面。

第一，从各类型指标数量和相对应的权重来看，五大评价体系中，科研能力指标所占的比例最高，达到42.6%，被赋值的权重比例甚至超过了指标比例，是教学水平指标的2.86倍，指标权重之比更是达到了4.25倍，反映出科研产出是各大评价体系最为关注的内容，也是大学评价的最重要指标。然而，五大评价体系基本上采用了国际可比的收录平台，科研能力在总指标权重和国际可比指标权重中分别占44.2%和36%，差异不大。但是，各大评价体系采用的数据库外的信息摄取显然是不够的，并不能客观体现大学真实的

更加关注教师教学水平的影响，荣誉称号与科研能力的关系更为紧密。

第三，全球可比指标被看作是一流大学的重要标准。从分析可以看出，全球可比指标个数总比例达到68.2%，全球可比指标权重总比例占到70.4%，ARWU和RCCSE的全球可比指标所占权重更是达到80%和88%。反映出国内外评价对一流大学建设存在普遍的共识。然而，其中也存在一些争议。比如五大评价体系基本上都采用国外的数据来源，似乎难以体现我国大学学术研究与教学的某些特殊性。同时，国外的评价体系基本以欧美国家的调查和语言为主，他们的主观性评价被纳入到全球性可比指标中，显然缺乏可信度，这也间接造成了我国的评价体系几乎以客观化标准为主的问题。

第四，教学水平受到关注。教学水平指标个数所占比例与国际化水平和声誉指标个数所占比例是一致的，但是无论从指标权重

表4 各类型指标个数(个)及权重(%)分布

指标类别	指标数量			指标权重		
	总指标数(个)	占总指标比例(%)	全球可比指标个数比例(%)	总权重	占总权重比例(%)	全球可比指标权重比例(%)
教学水平指标	7	14.9	4.3	52	10.4	3.4
科研能力指标	20	42.6	31.9	221	44.2	36
师资质量指标	4	8.5	6.4	52	10.4	8
知识转化收入指标	2	4.3	4.3	8.5	1.7	1.7
国际化水平指标	7	14.9	14.9	48.5	9.7	8.8
声誉指标	7	14.9	6.4	118	13.6	12.5

(10.4%)还是全球可比指标权重(3.4%)都表明,教学水平并没有得到足够多的关注,教学质量与教学指标间的客观化链接是困扰评价的主要原因。但是,这与国内外大学重视教学问题的现实似乎不符,并没有更精确地表现出国内外大学对于教学的投入和关注。事实上,国内外大学在教学中均投入了大量的资源,从近年来各大排行榜将本科生、研究生培养质量纳入评价指标体系就可看出其目的,甚至有的评价指标体系尝试将研究生培养与教师的科研产出相区分。就目前来说,各大评价体系并没有反映出大学间教学水平的差距,也难客观表征出大学教育质量的高低。可以预测的是,随着国内“双一流”建设的持续深化,教学质量必将得到更多的重视,从而实现高水平大学人员结构的良性循环。

第五,声誉指标在评价体系中的重要性正在逐渐提高。比如QS和THE的声誉指标权重近50%。声誉指标权重超过了教学水平和国际化水平指标权重,达到13.6%,全球可比指标权重达到12.5%,是教学水平权重的4倍,反映出声誉指标既是各大评价体系关注的内容,也是一流大学发展的内涵。RCCSE在最新完善的评价指标中创新性地加入了网络影响力,并作为声誉指标,客观反映出网络全球化时代人们对大学的关注方式的转变。

四、世界一流大学评价指标体系改进建议

从国外大学评价的主要目的而言,它们满足了社会民众对于政府和大学信息获取的需求,具有很强的价值偏好和引导作用,而从中国建设“双一流”大学的战略目标来说,世界一流大学评价指标体系应当作为大学质量保障和提升的重要组成。2017年教育部陈宝生部长曾提出“双一流”建设要实施“竞争优选、专家评选、政府比选、动态筛选”的原则,这四个原则涵盖了在评价的具体实施过程中的标准、内容、目的和主体等关键问题,其目的是科学而客观地制订符合世界一流大学标准和我国国情的指标体系,真正走出一条体现中国特色、展示中国风格、凸显中国气派的世界一流大学建设之路。基于此,笔者认为应当从以下五个方面改进和完善指标体系。

(一) 国际标准与中国标准的协调

目前,虽然研究考察的五大评价指标体系在指标设计、权重等方面存在显著差异,但基本上是以现在发展“最好的”国外大学标准作为参照,欧美的学术刊物和奖项作为依据,但是,从国际和国内各利益相关方的理解来看,似乎对于世界一流大学的标准并没有形成不可动摇的共识。完全照搬西方的标准对于引导我国大学发展

显然存在不利的影响,比如可能出现对潜在生源形成国外大学更好的价值引导,也可能带来外源性的文化自信缺陷。从各大排行榜的评价目的来看,潜在的大学生源和家长或消费者是评价机构认为的最大利益相关方,而缺乏对于政府作为大学投资者利益方的积极关照。比如在国际化水平因素指标上,由于历史、地域和语言的原因,欧美等国家国际化水平同学校的办学质量是否存在正相关仍值得商榷,ARWU和RCCSE并没有进行相应的指标赋予也说明国外标准的适应性问题。根据柏林原则,建议针对不同类型的大学开展不同方式的评价,中国的世界一流大学评价指标体系建设并不应该将国外大学的标准作为国内大学发展的根本性导向,需要兼顾国际标准与中国标准的统一,关照不同地域和文化属性对于大学质量和内涵的认识,尽可能地为未来大学发展留下想象的方向,能够在某种程度上反映大学未来可持续发展的动力。RCCSE在对国内大学的评价体系中,较好地关注了大学的既有存量和一定时间范围内的增量,也是构建中国的世界一流大学评价指标体系值得借鉴的经验。党的十九大报告明确强调,要实现高等教育的内涵式发展。所谓内涵式发展即审慎地面对实际利益的过度追逐,对现有西方主导的大学模式应理性地收敛心态。虽然我们从国外一流大学制度建设和外延式发展策略中获得了巨大的利益,然而,真正具有生命力的制度是基于国家、文化和学校土壤的结合,是大学自身难以被颠覆的活生生的原则和灵魂。

(二) 客观评价与质性评价的协调

借用评价技术的复杂化和评价指标的科学化来表达大学知识生产的规律是大学评价的前提假设,借用经济生产率的换算办法基本上是所有评价指标体系的共同做法。ARWU和RCCSE更青睐能够量化的评价指标,其中ARWU的所有指标几乎都指向了杰出人才和杰出成果的产出,“硬”指标在比较大学间科研产出量的方面优势明显,却难以体现出大学的历史价值、人文价值甚至是社会认可度,同时将过去的数据作为高低标准并不能体现当下教育质量的水平。竞争优选原则事实上回答了评价特征的发展方向,从大学排名发展的历史来看,主观性评价作为学术共同体内的评价从一开始便占据着很重要的地位,声誉影响几乎都由主观性评价完成,比如德国的CHE大学排名通过对250所高校的教师和学生定期调查,以实现“内部人的选择”声誉评价作为的独立数据。如邱均平教授所说:“没有科学的评价就没有科学的管理。”当然,科学的评价不等于量化的评价,上述五大评价体系都在试图体现自己的

科学性与合理性。大学在数千年的发展历程中，大学职能的增加直接带来了社会地位的提高，作为大学本身应该深思扩张的极限，严守大学固有的知识生产属性。因此，大学评价不能运用经济学和市场逻辑来看待其知识生产模式，更应该警惕和避免大学本身的科学逻辑弱化，客观评价与质性评价的兼顾与协调是构建科学的世界一流大学评价指标体系的主要方式。

(三) 大学质量与大学内在逻辑合理性的协调

大学评价的核心是大学质量还是大学其他因素呢？国内外学者普遍认同办学质量是大学排名的基本取向，或者作为大学质量保障的有机组成部分，双方互相促进共同提高大学的真实质量。然而，在大学质量的外在表现(指标)和大学内在逻辑合理性理解上却较难体现正相关。比如压力克斯·埃克尔指出：虽然不同评价机构使用了不同的评价标准，但是他们在好大学的评价上却体现出惊人的一致。再如，将就业作为大学质量的外在指标，事实上与大学内在逻辑也是不符的，学生就业既与学校、产业结构和经济等因素关联，也与个人意愿、社会价值观念和区域经济特征相关。大学质量与大学内在逻辑合理性的协调，事实上是要重新反思大学办学内涵、大学知识生产逻辑以及大学自身发展模式之间的关系、大学教育教学文化氛围形成等问题。如果说现代大学的主流模式是成为一流的研究型大学，那么在现有发展模式继续模仿前行的基础上，未来大型的发展模式又会是什么样呢？甚至于我们独有的大学模式应该是什么样？建设世界一流大学需要持久的内在动力，更需要大学对于本质、内涵等世俗目标之外的坚守。中国的世界一流大学评价指标体系构建需要深入思考评价体系与大学内在逻辑的相互统一，以期实现大学评价与大学办学的良性互动。

(四) 科研能力评价与育人功能评价的协调

从前面的数据分析结果看，科研能力仍然是各大评价机构积极认同的“一流”指标。然而，不可忽略的事实是，大学排名对于科学研究的价值偏好受到学者们的广泛诟病，这既与大学的根本目标认同有关，更与大学科研评价的正当性关系密切。比如威廉姆斯批评了ARWU对汤森路透数据库的全面依赖，以自然科学见长的数据库因为有限的期刊覆盖会导致严重的学科偏见。大学发展历史的整体文化表明，大学更像是一种“有组织的无政府状态”(organized anarchy)，教学和研究方面具有高度的个人自主性，依托大学人高度的自主和自治

实现其价值，某种价值性的偏好会导致大学功能性的失衡。我们认为，“双一流”建设应当继续坚持科学研究作为关键性评价标准不动摇，并以我国大学已经成熟、优秀的建设项目发展特征为基本导向；同时更应该秉承育人功能既是大学功能的历史起点，也是逻辑基础，更是大学之为大学的根本原因。令人欣喜的是，一些评价正在作出积极的改变，比如RCCSE将研究生科研成果不再计入学校科研能力指标，而列入研究生培养质量指标，事实上起到了强化大学育人功能的作用，这与我国近年来大力提高教学质量的举措不谋而合。

(五) 强、弱功利性评价的协调

大学具有强、弱功利性价值共在的特殊性，教育评价在体现其规范性和约束性、功利性价值的同时，也需要考量与发挥制度之善的弱功利性价值。所谓功利性评价的强弱讨论，并非基于价值判断孰优孰劣、孰是孰非的语境。事实上，功利性评价始终存在于大学及其功能的发展历史中，强功利性的评价价值在于激发大学及其发展逻辑中的共性和自觉性，而弱功利性的评价价值在于保护大学潜在的耦性和创造性。目前，科学而客观的评价方法已成为促进现代大学发展的重要内容，大学评价主要通过定量的科学评价指标体系而实现，“发表论文”不断挤压着大学原本存在的弱功利性价值生存空间，大学知识生产的特殊性已经被社会服务的功能压榨到极限，大学的强功利性价值被赋予前所未有的使命，而此种过度消耗的结果便是大学的神圣性几近丧失。近年来，大学出现了影响较为突出的学术道德不端、师德师风败坏等问题，大学的“象牙塔”形象不断受到侵蚀，如果说大学仍然是追求真理的神圣之地，那么，那些作为弱功利性价值的大学利益必须受到关注和保护。在强功利性价值大行其道的时候，必须保持大学主体性的清醒，呼唤和引导每一个知识分子的超越性理想，追求真理的学术信仰和拒绝器化(工具性)的德性节操，消弭其外在性和权威控制，转为回到大学追逐智慧和真理的返魅道路之上。可以认为，构建中国的世界一流大学评价指标体系应当也必须实现强、弱功利性评价的协调。

五、中国建设世界一流大学的启示

从各大排行榜排名来看，中国建设世界一流大学取得了令人欣喜的成绩，但是，我们还需要冷静地意识到评价只能说明大学在某些方面的成就。因此，在不断完善评价指标体系的基础上，建设中国的世界一流大学应当注意从关注一流的表象向关注一流的内涵转变、从关注一流的

量向关注一流的质量转变、从关注一流的科研向关注一流的学生转变。在此基础上，我们应当警惕唯数据与唯排名的思维、坚守求异作为大学本质、兼顾学术市场与大学变革。

(一)警惕唯数据与唯排名的思维

唯数据、唯排名的做法是将经济学、工业化的投入-产出式的企业化管理思维运用到大学管理之中。近年来，许多大学对本校的排名次序可谓殚精竭虑，想尽一切办法提高自身在各种排行榜中的位次，其中，最常见的也是最典型做法是迎合评价指标体系，以数据和排名论成败。通过分析评价体系中的某些指标，尤其是分析能够短时期内发生显著变化的指标，从而“不惜一切代价”提高排名，不惜挤占大学的其他发展目标、其他学科的发展空间。兹特和菲利亚热奥分析ARWU指标后指出：该指标体系具有难以消除的规模烙印和偏见，规模越大，排名越占优势，即规模大的大学在评价中更为有利。然而，事实却是大学之中杰出成就的产生只是少数人的作为，显然这与ARWU的规模偏好是相悖的。从大学自身的发展逻辑而言，大学质量属于价值范畴，即大学在某时期是否能够反映大学与人的需要的关联，大学是否能够为一流成果的出现提供更为契合的土壤，这才是最重要和最根本的东西。事实上，大学排名与大学质量之间的逻辑为非充要的关系。凡·布什曾告诫美国政府和大学：任何以牺牲其他学科为代价的计划都是愚蠢的。可喜的是，我们从教育部清除“三唯”的举措看到，相关部门已经作出积极的调整。

(二)坚守求异作为大学本质

历史进入现代社会以来，每个人都具备了成为“完整人”的可能性，也就是说，现代人因为外在世界的变化前所未有地具有独立感知世界的机缘，也因此，即使是关于“学以成人”的教育使命，也面临着前所未有的危机，教育难以继续给人以现成的价值和意义的确定性。显然，求同的做法不可避免抑制了大学的个性建设，一流大学各自的个性化特色在共性指标下遭到了遮蔽，这也是大学评价体系自产生之后便受到不断诟病的主要原因。我们在观察国外一流大学以及反思大学历史之后，能够得出两个明显的结论：一是愈加一流的大学，对于大学教授们的管理愈加宽松，并无严格的时间与成果指标限定，更注重积累之后的勃发；二是一流大学的超越性意义更为深远，即一流大学更为关注未曾发现、未被社会化的领域。我们从大学功能发展的逻辑来看，大学的每一次功能性延展和飞跃都是因为大

学自身求异的积极作为，因而求异才是大学得以生生不息的基本内涵。所谓求异并非说大学是独立于社会的存在，而是要从大学使命的自身扬弃出发，扎根国家和民族的文化需要，主动消解既有经验的缺陷，坚守大学办学的根本目的。

(三)学术市场与大学变革的兼顾

建设世界一流大学的命题说到底是如何提高大学的竞争力的问题。纵观公认的世界一流大学，其卓越的根本原因在于实现了学术市场和大学变革二者的兼顾与平衡。学术市场是每个时代大学都需要面临的现实，从大学功能的发展以及和经济、政治无法剥离的事实足以说明。到了今天，大学的学术市场需要面临的现实则是国际化的自由市场经济，大学需要足够的学术声誉换回发展的资源与竞争的砝码，因此，我们看到一流大学惯用的手段是争夺最优秀的教师和学生，事实上暗含着市场经济的运行机制。而作为大学变革的内部因素来讲，大学的历史及其在历史中积累的文化、传统，以及大学既有的职能、培养目标甚至管理模式，既是变革的主要对象，也是大学独有文化属性的见证。改革开放40年来，我国高等教育改革取得了快速的发展，但是深层的瓶颈问题仍然没有得到有效解决。因而，如何在大学体制改革的框架下，实现大学对学术市场竞争力的追逐和大学内部变革的协调兼顾，不但关乎是否能够建成世界一流大学的问题，更关乎我国大学未来发展的深远影响。

观照那些卓越大学的发展历史和深厚的文化积淀，既能感受到它们在历史变革中的果敢与坚毅，亦能体会到始终如一的稳定与保守，循序渐进似乎不足以涵盖它们的历史脉络，但是，我们却能看到循序渐进绝对是其最质性的品格。需要说明的是，基于五大评价指标体系关注的评价内容的简单类比，并不能充分反映出大学在知识创新等内涵建设中与服务国家经济社会重大战略中的实际贡献，也难以反映出大学排行差距产生的更深层原因。因而本文的目标只是为决策者和研究者提供一种客观观察的结果，为“双一流”背景下中国一流大学、一流学科的持续深入建设，以及构建具有中国特色的大学评价指标体系，提供一种有理论和实践意义的参考。

(原文刊载于《重庆大学学报(社会科学版)》2019年第5期)

(赵国栋，山西大学教育科学学院讲师、博士，山西太原 030006；马瑞敏，山西大学管理学院科学评价研究中心执行主任、副教授、博士，山西太原 030006)

中国特色世界一流大学 建设成效评价体系理论建构与实践验证

冯用军 赵 雪

一、中国特色世界一流大学第三方评价政策文本分析

中国特色世界一流大学第三方评价符合国家鼓励“管办评”分离的教育政策和教育行业第三方评估的世界潮流。当然，在高等教育领域推进专业评价和中介评估方面，我国也经历了一个较长的政策变迁过程。《国家中长期教育改革和发展规划纲要（2010-2020年）》提出要“鼓励专门机构和社会中介机构对高等学校学科、专业、课程等水平和质量进行评估。建立科学、规范的评估制度。探索与国际高水平教育评价机构合作，形成中国特色学校评价模式。建立高等学校质量年度报告发布制度”。

21世纪的中国，改革开放已经进入新时代，对人才的数量和质量的需求远胜往昔，建设若干所世界一流大学与一大批一流学科高水平大学的紧迫性更加凸显，这是中国大学的自身发展和中国经济社会的深化发展的规律使然与逻辑必然：第一，“双一流”大学持续培育的优质人力资源是支撑中国跻身世界强国、中华民族跻身世界优秀民族之林的核动力，也是支撑改革开放渡过深水区的最大红利；第二，中国高等教育经历了历史性变革取得了历史性成就，从精英到大众向普及化阶段坚实迈进，“双一流”大学正逐步成为社会轴心，特别是在大（数据）、智（能化）、移（动态）、云（计算）时代。

建设“双一流”大学是国之大计、党之大计、民族之大计，中国大学已经从“象牙塔”时代迈入“轴心”时代，大学的这种本质性的重大改变必然导致其角色的根本性变化，就是成为国家创新体系的重要部分和推动创新型国家建设的战略动能，是中国维系可持续进步、高质量发展的坚定不移的战略抉择。条条大道通罗马，世界一流大学的生长历史和发展现实均表明，每一所大学都有自身独特的成长理路和文化气质，它们走过的道路各具特色且无法复制。世界一流大学、世界一流学科高水平大学及其开展的世界一流本科教育，无论是在理想的状态下抑或在现实

的情境中，都是富于本国特色而多样化的，多样性是世界一流大学之所以一流的成功经验之一，这给中国开展“双一流”大学建设提供了一个重要启示。

二、中国特色世界一流大学的内涵界说与特征凝练

理想的大学、现实的大学都应该是多样性的，有时甚至是独一无二的，但在多样性中也能找寻通约性。换句话说，世界一流大学虽然有独特的DNA，但它们在高等教育的基本功能和现代大学的核心职能的履行上却有共通之处。这是对中国特色世界一流大学进行内涵构化与特征质化的前提。

在高等教育生态系统内，世界大学是多维度上的金字塔型，每一种类型、层次的大学都会有其佼佼者，它们在某一或某些领域领先一步或独步天下。所以，世界一流大学就是那些身处金字塔尖的研究型大学、专业型大学、应用型大学、技职型大学的聚合，它们把最有特色、最有优势、最有成就的学科群收入囊中，并千方百计维系它们在人才培养、科技研发、社会服务、文化传承和国际交流与合作等方面的领先地位和超前优势。

在党和国家战略布局中，中国大学是国家和平崛起和民族伟大复兴的利器，在“两个一百年”内，中国高等教育改革和发展的核心议题之一是“怎样建设（成）中国特色世界一流大学”，再造中国大学的世界影响、重构世界大学的中国格局。中国特色是中国大学的底色，华夏文明是中国大学的灵魂。中国特色世界一流大学是全方位凸显中华特色。

“我们要认真吸收世界上先进的办学治学经验，更要遵循教育规律，扎根中国大地办大学。”“双一流”战略规划中的一流大学建设高校（A、B类）、一流学科建设高校（C类）必须是中国人主导扎根中国大地通过自主创新建成的，绝不是直接照搬或简单复制发达国家的世界一流大学，但可以充分借鉴、吸收发达国家建设

世界一流大学、学科的经验教训，这就客观上需要对“什么是中国特色世界一流大学”的核心概念群进行内涵构化和特征凝练。

大学拥有相近的遗传基因和构成基质，但世界一流大学、世界一流学科高水平大学却是相同基因的不同排列组合，它们的气质、风格和文化独一无二且不可复制。促进社会发展和促进个体发展是大学的两大基本功能，培育人才、研发科技、服务社会、传承与创新文化、国际交流与合作则是大学的五大核心职能，世界大学在两大基本功能和五大核心职能的实现上都是通约的。

只是世界一流大学作为世界大学集团中的标杆，对它们的评价标准相对要高大上很多：一是大学理念的引领；二是毕业校友的杰出；三是科技研发的前沿；四是社会服务的卓越；五是传承与创新文化的坚守；六是国际交流与合作的优质高效。世界一流大学在多维教育生态系统内是相对的，在单维教育生态系统内是绝对的，它们的类型、层次和特征是多样化的，虽然每种类型、每个层次甚至每所大学都可以成为绝对一流（狭义相对）或相对一流（广义相对）大学，但跻身世界大学集团塔尖的大学永远都是极少数。

在多维高等教育生态系统内（世界高等教育组织约3万所），从认识论或政治论角度而言世界一流大学的数量、比例和生态位是相对稳定的，而从学术性、社会性双维度进行考察并参照全球主流大学排行榜，可以基本划定世界一流大学的分布比例与相对生态系统相位（表1）。

表1 中国特色世界一流大学分布比例和生态位次

办学分型				大学分层(%)	全球大学生态相位
综合研究型/应用型/技能型大学	行业特色研究型/应用型/技能型大学	专业研究型/应用型/技能型大学	职业技能型/应用型大学	世界顶尖大学 0.01	前 10 名
				世界一流/中国顶尖大学 0.05	前 11~30 名
				世界高水平/亚太顶尖/中国一流大学 0.1	前 31~50 名
				世界著名/亚太一流/中国高水平大学 0.5	前 51~100 名
				世界知名/亚太高水平/中国著名大学 1.0	前 101~200 名
				世界有名/亚太著名/中国知名大学 2.0	前 201~700 名
				亚太知名/中国有名/区域一流大学 5.0	前 701~3000 名
				亚太有名/省城一流大学 40.0	前 3001~10000 名
				市城一流大学 50.0	前 10001~

中国特色世界一流大学建设已历三年，按照国家“动态支持”的措施，2020年将迎来第一场

“淘汰赛”，所以2018年的中期建设绩效评估就至关重要。中央深改组审议通过的国务院《统筹推进世界一流大学和一流学科建设总体方案》，指明了中国大学的发展战略方向和分阶段发展目标，要求建立健全“双一流”大学建设绩效评价机制，积极采用科学性、公信度和权威性兼顾的第三方评价。

按照“双一流”战略的总体目标安排，在推进五大建设任务和五大改革任务、实施三大支持措施和两大保障措施“护法”下分三步走实现阶段性目标。如果这些战略目标能分步骤实现，那就意味着到2020年左右我国将有10余所大学稳居全球大学集团的5%，即稳居世界大学集团前百名；到2035年左右将有3-5所大学跻身全球大学集团的0.5%，即跻身世界大学集团前30名；到2050年左右将有1-2所大学雄踞全球大学集团的0.1%，即雄踞世界大学集团前10名。

但从全球趋势、中国国情和高教教情来看，我国高等教育战略目标落后于经济社会战略目标，必须实事求是地从理论上建构中国特色世界一流大学建设成效评价体系，通过“以评促建、以评促改、以评促管、评建结合、重在建设”加速我国“双一流”大学的建设步伐，为实现“中国梦”贡献最坚实的智慧。

三、中国特色世界一流大学建设成效评价体系的理论建构

发展是第一要务，人才是第一资源，创新是第一动力。中国特色世界一流大学必须有卓越的教学，打造一流学科、建设一流本科、做强一流专业、培育一流人才。一流教学是一流大学的底色，创新是一流大学的灵魂，一流大学的关键是培养一流人才，不仅要培养一流的科学家、工程师，还要培养一流的思想家、教育家。

世界一流大学和世界一流学科高水平大学都是在持续实现高等教育基本功能和核心职能上表现卓越的大学，也是对所在国家（地区）、民族和世界持续做出伟大贡献的伟大大学。世界大学虽数量庞大，但伟大大学如同伟大人物一样屈指可数，伟大人物都有相似的卓越品质，伟大大学亦是如此。所以，世界一流大学的共性品质就是评价中国特色世界一流大学建设成效的参照标准和判别准绳，也是研发、构建中国特色世界一流大学评价体系的实然准则和理论基石。

“双一流”战略是百年大计，事关中国的千年发展，扎根中国大地办大学，既然第一所世界

一流大学必须由中国人主导在中国大地上建成，那么评价“双一流”大学建设成效的评价体系也必须是由中国人主导在中国大地上构建。那些外国主导的大学排名或那些中国人简单照搬外国排名的学术排名，既未按照“世界眼光、国际标准、中国特色、高点站位”来研发和运行，也没有多少中国特色和中国元素（基本是THES或QS等的翻版“舶来品”或“复制品”），这为中国大学人提供了千载难逢的历史机缘。

在“双一流”战略背景下，摆在中国大学评价主体面前最大的新时代政治任务，就是建构中国特色鲜明、具有世界眼光、与国际无缝接轨、高点站位的世界一流大学评价体系。大学评价不应是一门生意而应是一门计量科学，由中国教育家联合中国科学家主导并借鉴国际大学排名机构的经验研发、制定“双一流”大学建设绩效评价机制和体系，是中国世界地位和全球影响力在高等教育发展等方面的客观要求，也反映了由中国主导制定世界大学排名标准、发起“一带一路”高等教育繁荣倡议、构建全球大学命运共同体等的现实需要。

教育主权是国家主权的重要方面，教育评价权是教育主权的有机组成部分。全球大学排名权是一个国家、一个民族世界话语权的重要体现，也是一个国家、一个民族展示高等教育软实力和核心竞争力的重大平台，更是中国教育战略家、大学排名专家引领世界高等教育评价事务的重大舞台。

维护国家教育评价权是一个国家和其他组织开展平等教育交流的根本前提，本国教育评价组织对本国大学进行评价具有优先权、对他国大学评价具有选择权。中国可以主导的全球大学排名的数量与质量，是中国人全球大学排名领域话语权与影响力的“晴雨表”。大学排名诞生35年来，发布过大学排行榜的机构或个人当以百计，但真正拥有世界影响且持续发布排行榜的却以个数，对它们的“核心指标”进行“合并同类项”后发现：学生质量（师生数量、校友成就）、科技贡献（成果数量、成果质量）与国际影响力（学术声誉、媒体影响）是评价大学的三大核心标准。

为跟踪评价、动态监测、科学预判中国特色世界一流大学建设成效，“双一流”大学评价体系建构必须融汇中国特色、中国文化和中国元素，站在第三方评估立场，展示中国风格、中国

气派和中国精神。以高等教育的两大基本功能和现代大学的五大核心职能等为基础，借鉴国内外高等教育评估和现代大学评价的经验教训，科学地构建起中国特色世界一流大学评价指标体系，该指标体系包括2个一级指标、6个二级指标、9个三级指标、12个四级指标以及50个五级指标观测点。

该评价指标体系是中国自主创新、集成创新和体制机制创新的成果，是在30余年大学评价科研、设计、验证、试点和运行经验的基础上，投入大量精力和智力，充分借鉴国际第三代大学评价技术先进理念，汲取诸多大学评价机构的教训反馈，遵循大学排名国际专家组

表2 中国特色世界一流大学建设成效评价体系

元指标	一级指标	二级指标	指标权重	指标侧重	指标属性
人才培养 (教育教学/ 文化传创) T&CI	教学水平	教学质量	0.08	中国特色	定性/历时数据
	教学水平	校友质量	0.12	世界标准	定量/历时数据
	师资队伍	教师质量	0.09	世界标准	定量/历时数据
	培养基地	学科实力	0.074	中国特色	定性/历时数据
科技研发 (理论研究/ 技术研发) S&TI	学术项目	科技项目	0.07	中国特色	定量/历时数据
	学术平台	科技平台	0.07	中国特色	定量/历时数据
	学术成果	科技成果	0.1	世界标准	定量/历时数据
	学术奖励	科技奖励	0.095	世界标准	定量/历时数据
社会影响 (社会服务/ 文化传创) S&CI	社会声誉	校友捐赠	0.04	世界标准	定量/历时数据
	社会声誉	成果转化	0.055	中国特色	定性/历时数据
	媒体影响	国内媒体	0.016	中国特色	定性/历时数据
	媒体影响	国际媒体	0.016	世界标准	定性/历时数据
国际声誉 (国际交流 与合作) I&CI	国际师生	境外师生	0.04	世界标准	定量/历时数据
	国际平台	跨境平台	0.05	中国特色	定量/历时数据
	国际学术	学术影响	0.044	世界标准	定量/历时数据
	国际排名	世界排名	0.04	世界标准	定量/历时数据

(International Ranking Expert Group, IREG) 提出并倡导的“高等教育机构排名的柏林原则”(Berlin Principles on Ranking of Higher Education Institutions, BPRHEI)，构建起了按中国特色、世界标准研发设计的第四代大学评价技术(大学排名4.0)，即提出了破解世界大学评价难题的中国智慧和方案——兼和世界眼光、国际视野和中国特色、华夏风格的世界一流大学建设成效动态监测与多维评价体系(表2)，对“双一流”高校建设成效进行动态监测、跟踪评估并提供质量保障、遴选建议。

从理论上讲，这套指标体系的最终目的是构建“中国特色世界一流大学评价指数”(The Chinese Top Universities Index, CTUI)，

其中未来优化迭代的方向是构建元指标（促进社会发展、促进个体成长）、一级指标（培育人才、研发科技、服务社会、国际交合、传创文化、综合声誉等）、二级指标（生源质量、师资队伍、培养平台、校友质量；科技项目、科技平台、科技成果、科技奖励；校办企业、专利转化、咨询报告、社会捐赠；国际化教育、国际化科研、国际化合作[国际合作联合实验室]、国际化奖励；文化传创平台[国际华文教育基地学校]、文化传创项目[孔子学院/课堂]、文化传承成效[全国文明校园、教育部高校校园文化建设

2017年12月31日）与标杆大学进行了实践验证，经过数据清洗和标准化归一处理，其建设进展和国际排名情况如表3。

分析表3可以发现，中国特色世界一流大学已经取得了比较明显的建设成效，有了一个良好开局，为有计划建成若干所世界一流大学和一批世界一流学科高水平大学奠定了良好基础。具体而言，清华、北大已挺进世界大学集团的前0.5%，是目前中国大陆最接近世界一流大学行列的世界一流学科高水平大学，是“双一流”大学集团率先实现2020年阶段性目标的领头羊；复旦、上交大、浙大、中科大、国防科大、南大已挺进世界大学集团的前5%，是目前中国大陆最接近世界一流学科高水平大学行列的世界著名、亚太一流大学，是建设中国特色世界一流大学的主力军；中大、哈工大、北师大、武大、华科大、同济大学、西安交大、南开大学、厦大、东南大学跻身全球大学集团的前1%，是世界知名、亚太一流学科高水平大学，是建设中国特色世界一流学科高水平大学的生力军。新时代教育评价人的新使命是增强中国在世界教育评价领域的话语权，为世界大学评价提供中国智慧和中国特色方案，为中国建设世界教育中心服务。

表3 中国特色世界一流大学建设成效评价

排序	校名	国际排名	国家定位	排序	校名	国际排名	国家定位
标杆大学	哈佛大学	1	—	标杆大学	麻省理工学院	2	—
1	清华大学	28	A	22	山东大学	232	A
2	北京大学	30	A	23	天津大学	234	A
3	复旦大学	59	A	24	大连理工大学	237	A
4	上海交通大学	64	A	25	中国人民大学	238	A
5	浙江大学	68	A	26	北京航空航天大学	240	A
6	中国科学技术大学	70	A	27	北京理工大学	259	A
7	国防科技大学	93	A	28	华东师范大学	265	A
8	南京大学	98	A	29	中南大学	278	A
9	中山大学	126	A	30	兰州大学	280	A
10	哈尔滨工业大学	145	A	31	中国农业大学	284	A
11	北京师范大学	155	A	32	湖南大学	301	B
12	武汉大学	155	A	33	电子科技大学	326	A
13	华中科技大学	167	A	34	重庆大学	361	A
14	同济大学	180	A	35	东北大学	383	B
15	西安交通大学	182	A	36	西北工业大学	407	A
16	南开大学	184	A	37	中国海洋大学	408	A
17	厦门大学	199	A	38	郑州大学	517	B
18	东南大学	200	A	39	云南大学	561	B
19	华南理工大学	218	A	40	西北农林科技大学	623	B
20	四川大学	230	A	41	新疆大学	674	B
21	吉林大学	231	A	42	中央民族大学	712	A

优秀成果]；媒体影响、国际声望、世界排名）和四级指标观测点既象限离散又耦合共轭的大学生态评价系统，实现对“双一流”大学建设全要素的绿色生态评价。

四、中国特色世界一流大学建设成效评价的实践验证

为了检验中国特色世界一流大学建设成效评价体系的信效度，遵循科学、独立、客观、透明、公正的原则，采集深度反映42所一流大学建设高校（A+B类）五大职能的核心数据（截至

教育强国，评价先行。教育评价工作者应努力建构中国特色鲜明、世界标准接轨的第四代评价模式，即多元主体、协商平等的评价思想，中外融汇、智能进化的评价制度，贡献优先、兼顾公平的评价原则，立体监测、注重诊断的评价过程。多元与融合是新时代评价模式的核心特征。在“双一流”评价实践中，要主动借鉴和吸收国外评价发展性的价值取向、投入产出的内容指向、认证及专业机构的制度保障等主要经验，但绝不能丢掉中国特色、绝不可缺少中国元素、绝不要淡化中国文化。“双一流”大学建设成效是衡量中国发展水平与发展潜力的重大标志，站在第三方立场、以贡献优先、独立地对一流大学建设高校、一流学科建设高校的建设成效进行跟踪监测、动态评估与科学遴选、分类指导是党的要求、国家需求和人民需要。

（原文刊载于《江苏高教》2019年第1期）

（冯用军，北京师范大学教育学博士后，北京100875；赵雪，唐山师范学院讲师、硕士，河北唐山063015）