

文章编号: 2095-1663(2023)01-0063-09

DOI: 10.19834/j.cnki.yjsjy2011.2023.01.09

“双一流”建设中的第三方评价

——从“评价崇拜”到“评价赋能”

崔育宝, 李金龙, 张淑林

(中国科学技术大学 公共事务学院, 合肥 230026)

摘要: 第三方评价以其独立性、专业性、公正性等特点在“双一流”建设中得到高度认可, 成为“双一流”建设评价的重要组成部分。然而, 当前的第三方评价存在着世界一流品质体现度低、社会贡献评价导向性弱、科学性与公正性难保障、促“建”作用发挥不足、支撑动态调整有缺陷等问题, 并不能承受其重。未来, 在“双一流”建设推进过程中, 对待第三方评价的态度应从“评价崇拜”向“评价赋能”进行转变。具体地, 要从“存量”和“增量”两个层面推动第三方评价更好地服务和促进“双一流”建设, 即一方面要在“双一流”建设中正确认识和规范使用现有第三方评价, 另一方面要鼓励发展符合“双一流”建设导向的新型第三方评价。

关键词: 第三方评价; “双一流”建设; 元评估

中图分类号: G643

文献标识码: A

2015年10月, 国务院印发《统筹推进世界一流大学和一流学科建设总体方案》, 提出要“积极采用第三方评价, 提高科学性和公信力”。随后出台的系列“双一流”政策也都旗帜鲜明地强调第三方评价应在“双一流”建设的遴选评价和成效评价中发挥重要作用。然而, 开展“双一流”建设评价本身不是目的, 目的在于促进“双一流”建设。作为“双一流”建设评价的重要组成部分, 第三方评价应从被“双一流”建设“崇拜”向为“双一流”建设“赋能”转变, 以更好地促进“双一流”建设战略目标实现。为此, 本文将从“双一流”建设评价的现实诉求出发, 分析第三方评价在服务“双一流”建设中存在的不足, 探究第三方评价促进“双一流”建设高质量发展的路径。

一、高等教育第三方评价及其分类

学术界对高等教育第三方评价的界定有两种观

点: 一是强调组织独立性是保证评价结果公正的起点, 认为评价主体的独立性是识别高等教育第三方评价的必要条件, 即从事高等教育第三方评价的组织必须具有独立于政府和高校之外的法人地位^[1]; 二是认为评价活动是否具有公正性的关键不在于“谁来评”, 而在于它的评价标准及评价过程是否具有第三方特征(公正性)^[2]。从评价实践来看, 即便是符合组织独立性特征的第三方评价也往往会在自利动机和谋利行为影响下成为“失效的第三方”, 被诟病为伪独立、伪专业、伪客观^[3]。也有学者指出在中国现实情境下, 应从“依附式自主”视角重新审视高等教育评价中第三方组织与政府的关系^[4]。鉴于“第三方”主体身份的复杂性以及“教育评价”活动的特殊性, 本文认为从广义角度界定高等教育第三方评价更具现实意义, 即只要使用的是具有第三方特征(公正性)的标准, 不论是谁都可以施行“第三方评价”^[2]。具体地, 高等教育第三方评价是指由符合相

收稿日期: 2022-11-25

作者简介: 崔育宝(1989—), 男, 山西高平人, 中国科学技术大学公共事务学院特任副研究员, 博士。

李金龙(1987—), 男, 河北邯郸人, 中国科学技术大学马克思主义学院讲师, 博士。

张淑林(1956—), 女, 安徽蒙城人, 中国科学技术大学原副校长, 研究员; 中国学位与研究生教育学会副会长。

基金项目: 教育部人文社会科学研究一般项目青年基金“第三方评价元评估及其在‘双一流’建设中的运用研究”(20YJC880009)

应资质要求的第三方机构或非当事方组织按照公正标准和科学程序,对高等教育各类活动进行评价并对结果负责的活动。

鉴于“双一流”建设的建设目标及其在遴选评价与成效评价中的相关评价标准^[5],本文梳理了部分具有代表性的高等教育第三方评价(表1)。

表1 代表性的高等教育第三方评价

评价机构/组织实施者	评价名称
《美国新闻与世界报道》	全球最好大学排名、 全球最好大学排名—学科
《泰晤士高等教育》	世界大学排名、世界大学排名—学科
QS公司	世界大学排名、世界大学排名—学科
科睿唯安公司	ESI大学排名、ESI学科 排名、高被引作者
爱思唯尔公司	中国高被引作者
上海软科公司	世界大学学术排名、世界一流学科排名、 中国大学排名、中国最好学科排名
教育部学位中心	学科评估
国务院	国家科学技术奖
教育部	高校人文社科优秀成果奖
高校	院系/学科国际评估

其中,学科评估是教育部学位与研究生教育发展中心(以下简称学位中心)以第三方评价方式开展的评价活动,在评价组织与实施,评价指标体系设计、评价方法等方面尽可能地保证了评价工作的专业性与公正性^[6]。国家科学技术奖、教育部高校人文社科优秀成果奖等评价活动虽然是由政府组织实施的,但其都会通过制定相应的章程或办法来保障评价组织与评价过程的科学性、公正性,且依据章程或相关办法成立的奖励委员会、学科评议小组在组织层面与高校、政府间保持了一定的独立性。院系/学科国际评估是由高校自主聘请国际知名学者组成评价小组展开的诊断式评价,其目的在于发现问题进而实施变革^[7]。从评价标准及评价过程是否具有第三方特征(公正性)来看,院系/学科国际评估往往是在无外部考核压力以及学校内部具有强烈变革意愿的牵引下开展的^[7],遴选的专家组成员往往具有较强的学术影响力和问题洞察力,可以保障评价组织的专业性及评价结果的公正性;从第三方评价理论与实践来看,第三方评价可分为委托第三方评价和独立第三方评价,委托第三方评价既可以是委托第三方机构,也可以是组织专家成立第三方评估组^[8]。国际经验表明,纯粹意义上的第三方评价只是一种理想化追求,为避免“外行评价内行”风险,先行国家的政府绩效评价往往只是“吸纳”第三方评价专家、科学家参与,而非简单地交由第三方机构去实

施^[9];从推进“双一流”建设来看,开展院系/学科国际评估有利于加强高校和学科与国际同行的对话与交流,增强内生动力,助力高校“双一流”建设,促进“双一流”建设战略目标实现。综上,本文将高校自主实施的院系/学科国际评估也纳入到了与“双一流”建设相关的高等教育第三方评价之中。

二、“双一流”建设评价的现实诉求

“双一流”建设既是我国追求建设高等教育强国的政府行为,也是关乎我国高等教育高质量发展的教育工程。“双一流”建设评价理应在服务政府绩效管理、提升高等教育发展质量、激发高校办学活力等方面作出积极回应。

(一)监测“双一流”建设成效

“双一流”建设本质上是政府在高等教育领域实施的国家战略行为。它既在政策制定层面提出了“双一流”建设的目标与任务,也在实施过程中对“双一流”建设高校给予了声誉、政策、资金等资源支持。对“双一流”建设的总体目标实现情况、建设任务与改革任务执行情况、财政资金合规使用情况进行监测与评价,既是政府绩效评价与管理工作的基本要素,也是推进高等教育治理现代化的必然要求。

在“双一流”建设成效监测与评价方面,2020年底印发的《“双一流”建设成效评价办法(试行)》(以下简称《成效评价办法》)指出,“双一流”建设成效评价应由大学整体建设评价和学科建设评价两部分组成,按“前置维度+核心维度+评价视角”方式布局考核内容:前置维度考察建设高校“加强党的全面领导与治理体系改革成效”;核心性维度反映“双一流”建设总体方案的五大建设任务和五大改革任务;评价视角考察整体发展水平、成长提升程度、可持续发展能力等^[10]。同时,考虑到“双一流”建设的目标,该办法还提出要科学合理确定相关领域的世界一流标杆,对建设高校和学科在全球同类院校相关可比领域的表现、影响力、发展潜力等进行综合考察。

(二)推动“双一流”建设发展

“双一流”建设评价同时也是一种教育评价行为。教育评价天然具有导向功能和改进功能:导向功能体现为价值判断具有主体性,即“有什么样的评价指挥棒,就有什么样的办学导向”;改进功能则体现为教育评价的目的在于达到“教育价值增值”^[11]。对“双一流”建设评价而言,就是要在评价内容、评价标准、评价结果的设计与运用上符合“双一流”建设的战略导向,促进“双一流”建设目标实现。

从评价导向来看,“双一流”建设评价的核心导向就在于“中国特色、世界一流”。其中,“世界一流”是指要引导大学与学科建设追求卓越,在国际可比性领域进入世界一流行列或前列;“中国特色”是指要引导大学与学科建设服务社会需求与聚焦社会贡献。从评价目的来看,正如美国教育评价专家斯塔弗尔比姆所言,评价最重要的目的不是为了证明,而是为了改进。因此,“双一流”建设评价的目的还在于通过评价识别问题、揭示原因、实施变革,促进大学与学科发展水平持续提升,实现建成世界一流大学、世界一流学科和高等教育强国的目标。

(三)服务“双一流”动态调整

建立动态调整机制,打破身份固化,不搞终身制是“双一流”建设在高等教育重点建设上的重大制度创新,有助于调动和激发高校办学的积极性与发展活力。“双一流”建设的动态调整至少包含三方面涵义:建设高校资格的新增与取消、建设学科资格的新增与取消、建设经费的加大与减少^{[12]44}。其中,高校建设资格、学科建设资格、建设经费是“双一流”建设动态调整的内容;新增与取消、加大与减少是“双一流”建设动态调整的措施;相关评价结果是“双一流”建设动态调整的重要依据,其科学性、公正性将直接决定“双一流”建设动态调整机制运行的有效性。

在“双一流”建设的动态调整中,建设经费的加大与减少只涉及建设高校与建设学科,建设资格的新增与取消则会涉及建设高校与候选高校、建设学科与候选学科。从“双一流”建设的政策表述和具体实践来看,建设资格的新增依据的是建设周期前基于一定标准的遴选评价结果,建设资格的取消依据的是建设周期期末的成效评价结果,建设资金的加大与减少依据的也主要是建设周期期末的绩效表现与成效评价结果。遴选评价与成效评价遵循何种评价标准,如何构建遴选评价结果、成效评价结果与“双一流”建设资格与建设经费的动态调整关系,是“双一流”建设评价面临的重大理论和实践难题。

三、当前第三方评价不能承受之重

不可否认,现有诸多第三方评价自诞生之日起便不是以服务“双一流”建设为导向的,其无法承受“双一流”建设评价所要承担的重要使命本就是一个无可辩驳地的事实。但当第三方评价作为衡量高等教育质量的重要方式要在“双一流”建设的遴选评价与绩效评价中发挥重要作用时,我们有必要分析这些被“双一流”建设评价所采信或可能采信的第三方

评价(如表1所示)能否有效承载起或如何无法承载起“双一流”建设评价所要承担的重要使命。

(一)世界一流品质体现度低

“双一流”建设的主要目标在于推动我国越来越多的大学和学科进入世界一流行列或前列,比较直观的理解便是我国有越来越多的大学和学科能够在全球范围内可比的领域中处在靠前的位置。上海交通大学于2003年首次发布世界大学学术排名(上海软科世界大学学术排名的前身),其目的就在于定量分析我国大学在世界大学体系中的位置,找出与世界一流大学的主要差距。随后,针对全球范围内大学与学科的排名如雨后春笋般发展起来,其中最具广泛影响力的包含上海软科公司(以下简称上海软科)、《美国新闻与世界报道》(以下简称 U. S. News)、《泰晤士高等教育》(以下简称 THE)、QS 公司(以下简称 QS)发布的大学/学科排名和科睿唯安公司(以下简称科睿唯安)基于 ESI 分析工具发布的大学/学科排名等。

总体上,这些世界大学/学科排名在方法论和指标选取上并非完全以世界一流品质为导向。大学/学科排名方法本质上是一种多指标综合评价方法,评价指标与指标权重是影响排名结果的两大关键因素。仅从评价指标设计来看,大学/学科排名中的评价指标大体可以进行量与质的分类:代表量的指标包含师生比、国际师生比、教师博士学位获得比例、研究投入、发表论文总量等,这类指标的表现往往可以通过资源投入在短时间内得以提升(如果要刻意追求的话);代表质的指标包含获得重要科研奖项的数量、发表在顶级期刊的论文数、考虑论文引用的评价指标以及一些学术声誉调查等,这类指标数值的提升往往会对大学与学科的资源投入、人才集聚、制度建设、学术沉淀等提出更高要求。考虑两种极端情况:如果全部用量的指标进行评价,那么一些世界公认的一流大学与学科或许无法居于排名前列,排名的公正性就会存在争议,难以获得社会认可;如果全部使用质的指标进行评价,那么就无法吸引更多的大学与学科参与世界大学/学科排名的“游戏”中(很多高校会是零分),久而久之就会成为少数大学的狂欢。从扩大影响力来看,大学/学科排名如果参与规模小且位次长期稳定不变的话,就无法为高校与社会公众带来“新闻事件”与“谈资”,其社会关注度也会随之走弱。因此,世界大学/学科排名评价指标体系中往往兼具量的指标和质的指标,它们只是常规性的一般评价而非卓越导向的质量评价,它们更偏向于市场而非学术。此外,即便一些代表质的

指标,其表现也可以在实践过程中“以量取胜”,使质的评价在无形中被缩小。如 THE、U. S. News、ESI 中使用的论文引用次数指标, QS 的师均引用次数指标,科睿唯安的高被引作者评选等,既可以通过发表高水平论文,以篇均引用次数提升带动总论文引用次数提升,也可以通过扩大发表论文数量提高总论文引用次数和师均引用次数。根据最新的 ESI 最新数据(2022 年 6 月),中国大陆科教机构发表论文的总被引次数和热点论文数位居世界第二,但篇均引用次数在论文总引用次数排名前十的国家与地区中为倒数第一,在论文总引用次数排名前十的国家与地区中也仅高于印度、巴西和伊朗。

(二)社会贡献评价导向性弱

服务社会需求,聚焦社会贡献是“双一流”建设的重要导向。大学与学科的社会贡献是其在履行自身职能过程中而对人类社会的政治、经济、科技、教育、文化、军事等发展产生的积极作用,其社会贡献评价可从人才培养贡献、科学研究贡献、社会服务贡献、文化传承与创新贡献等方面展开^[13]。从建设中国特色的世界一流大学与世界一流学科角度来看,“双一流”建设背景下大学与学科的社会贡献评价应注意以下几点:一是要注重产出评价。大学与学科的社会贡献属于教育外部关系范畴,反映的是教育活动与社会发展的单一向度关系,因此应注重从产出角度进行社会贡献评价,而非如针对教育活动自身评价那样混合使用投入、过程、产出等多种指标类型;二是要体现与国家需求的契合度。教育外部关系规律决定了教育必须为一定社会的政治、经济、文化科学服务^[14],针对我国大学与学科的社会贡献评价自然应在评价标准上体现产出与国家需求的契合度,彰显“中国特色”。三是要体现高标准要求。“双一流”建设高校与学科的发展基础好、资源投入多、社会期望大,对其社会贡献的评价应高标准、严要求,以起到促进其追求卓越、迈向“世界一流”的作用。

结合以上三点讨论可以发现,目前的第三方评价存在社会贡献评价导向不足的问题:一是较多使用投入型、过程型评价指标。如师生比、机构收入、研究投入、科研项目获取、科研平台建设等;二是评价标准与国家需求的契合度不高。一方面,世界大学/学科排名评价以科研评价为主,并主要是在西方学术话语体系下考察英文期刊论文发表情况和展开学术声誉调查,对与区域发展密切相关的工程应用学科、人文与社会学科的贡献评价重视不足。另一方面,世界大学/学科排名评价还存在重视最新成

就、忽视历史贡献的问题。世界大学/学科排名评价在对期刊论文发表情况考察时,往往将数据采集时间限定在近 1 年、近 5 年、近 10 年等时间范围,这一做法或将引导我国高校持续不断地在他国科研基础上做短、平、快的科技创新,而忽略对仍处于追赶阶段的、不易发表论文的科技领域进行攻坚克难;三是卓越评价导向不足。如前所述,所有大学/学科排名评价中的科研评价指标都会兼具量的指标与质的指标,并非坚持卓越导向。此外,产业收入、雇主声誉调查等所谓社会服务类指标的得分高低显然不能成为区分一所大学是否为世界一流大学的核心要素。

(三)科学性与公正性难保障

评价本质上是一种价值判断。不同的评价主体往往会采取不同的评价策略,进而对同一客体的评价形成不同的评价结论。面对日益繁多的评价活动,有必要通过开展元评估来确保原评价结论的合理性与准确性。下面主要从评价目的、评价内容与标准、信息收集与处理、结果公布等四方面对第三方评价的科学性与公正性展开分析。

评价目的。评价目的决定评价内容与标准,评价组织类型则会从根本上影响到评价目的。政府是高等教育事业的举办者,在高等教育资源配置中起基础性作用,其开展评价的主要目的在于实现对评价客体的监测、激励与管理等。企业是以营利为目的的经济组织,其开展第三方评价的目的自然就是服务相关经济活动,包括为学生与家长提供留学咨询服务、为相关高校提供排名咨询服务,为高校绩效评价与提升提供咨询等。ESI 排名是学术信息出版公司 Institute For Scientific Information(美国科学信息研究所,ISI)于 2001 年推出的衡量科学研究绩效、跟踪科学发展趋势的基本分析评价工具 Essential Science Indicators(基本科学指标,ESI)中的一项分析指标,但 ISI(历经多次资产重组后成为当前的科睿唯安公司)一直都是一家以市场盈利为目的的商业公司,其所开展的影响因子评价活动本质上是一种以学术为名的营销手段^[15]。高校是高等教育的办学主体,其自主开展第三方评价的目的就在于发挥评价的诊断性功能、实现以评促建,如学科/院系国际评估等。

评价内容与标准。现有高等教育第三方评价大体可以分为单项评价和综合评价。单项评价是从单一维度来反映大学与学科的建设成效,如 ESI 排名、高被引作者评价和重要奖项评选等。综合评价是从多个维度来综合反映大学与学科的建设成效,如不同机构发布的大学/学科排名等。相较单项评

价,以排名评价为代表的综合评价往往会诱使社会公众认为其所呈现的结果就是大学与学科的真实状况。事实上,大学/学科排名评价是从构成论视角出发,将大学与学科分割为人才培养、科学研究、社会服务、国际交流与合作等要素,以可测量、可量化、易获取等为原则选取评价指标展开评价。总体而言,现有大学/排名评价在评价内容和标准上存在如下问题:(1)评价指标与评价内容错位。如以师生比指标评价教学质量,以英文论文发表及其引用情况、期刊等级评价科学研究水平等。(2)重量轻质,非以卓越为导向。这是引入更多“玩家”进入排名游戏的必要手段。(3)一把尺子量天下,忽视分类评价。以研究型为标准进行评价,忽视各大学与学科的发展特色、发展阶段、所处的经济社会环境及其发挥的重要贡献。此外,针对大学的综合评价存在着重视理工类高校,轻视人文类高校的倾向,针对学科的评价存在以期刊定学科的情况,且学科分类体系与我国的学科分类体系并不相符等。

信息收集与处理。开展大学/学科排名评价的机构大多为商业公司,相关评价信息在收集与处理过程中如果缺乏必要的审查与监督,就容易被商业化污染。尤其对于非公开渠道获取的评价信息而言,评价机构往往可以采取合法的方式来影响评价结论,造成评价结果的失真。如开展声誉调查时,采取不同的样本采集策略就会得到不同的调查结果。

结果公布。总体而言,公布的评价结果信息越多,越能增强社会公众对评价客体的了解,有利于相关大学与学科发现差距与问题;同时,也越能反映出评价机构在信息收集与处理过程中的公正性,其结果经得起社会公众的质疑。上海软科、THE、QS、U. S. News 的世界大学排名都会公布各大学的总排名次序和总百分制得分,但在更细化的信息公布上有所不同:上海软科、QS 的世界大学排名公布了各大学各评价指标的百分制得分;THE 世界大学排名虽然设计了更为丰富的评价指标体系,但在结果公布上仅呈现了一级评价指标的百分制得分;U. S. News 世界大学排名仅公布各大学在各评价指标上的排序。虽然百分制得分的计算方式也会隐藏更多真实信息,但排序相较得分而言,更存在失之毫厘、差之千里的问题。值得关注的是,科睿唯安的 ESI 排名的数据源自其旗下产品 Web of Science 核心合集数据库。该数据库对社会开放,ESI 中的排名数据也就可供社会公众进行核查并溯源到每一条源数据,一定程度上确保了 ESI 排名评价的准确性与公正性。但 ESI 排名以期刊定学科,仅以论文引用衡

量学科水平的做法显然也是不科学的。

(四)促“建”作用发挥不足

评价的目的在于实现改进。但从实践来看,相关的第三方评价在帮助高校与学科识别问题与差距,乃至提出变革建议上能够发挥的作用还很有限。

一是源于评价内容的影响。评价活动的促“建”作用发挥的前提在于能够准确了解现状和识别问题,而准确了解现状和识别问题的前提在于保证评价内容的科学性与完整性。如前所述,目前绝大多数第三方评价在指标设计与数据来源上往往难以保障科学性与公正性,且存在对“中国特色、世界一流”评价导向重视不足的问题。此外,现有第三方评价在评价内容上局限于“五大建设任务”层面,忽视对“五大改革任务”相关内容的评。即往往从构成论视角出发,将大学与学科分割为人才培养、科学研究、社会服务、国际交流与合作等要素,重点选取可测量、易比较的代表性指标进行评价,忽视从生成论视角出发,围绕生成世界一流大学和世界一流学科的学术精神、组织氛围、治理体系等展开评价。

二是源于评价类型的影响。目前针对大学与学科的评价多为终结性评价,注重从构成论视角设计评价内容,在结果呈现上仅提供反映大学与学科建设水平的排名或得分,不负责分析排名或得分高低的原因,也不关注大学与学科建设过程中采取相关举措的有效性、积累的成功经验和面临的管理弊端等^[16]。以学科/院系国际评估为代表的诊断性评价注重从生成论视角设计评价内容,系统分析相关学术组织的战略规划、优势和劣势、面临的危机与挑战,进而提出建设性策略和变革思路^[7]。显然,诊断性评价更有利于发挥评价工作的促“建”作用。但吊诡的是,学科/院系国际评估并未在高等教育界引起足够重视,以排名评价为代表的终结性评价的社会影响力却在日益增强。

三是源于评价效用的影响。无论是社会中介机构开展的还是由政府主导实施的第三方评价,本质上都是竞争性评价、问责性评价。一方面,这些第三方评价以简单粗暴的排名方式呈现评价结果,掩盖了大学与学科的整体状态与潜在内涵,无法准确反映真实状况,导致其指导效果大幅削减^[17]。另一方面,这些第三方评价结果绑定了太多利益因素,促使标榜学术自主、崇尚创新创造、追求个性化办学的高校不得不自觉地按照评价指标的要求规范自己的办学行为,将评价指标异化为学术目标,过多关注那些可能会对可测量绩效产出有积极影响的活动,忽视与可测量价值无直接相关的社会性、情感性和道德

性的发展活动^[18]。此外,部分高校还采取信息伪饰、材料拼凑甚至是与评价机构交易的方式获得排名次序的提升,严重破坏高等教育系统的健康生态。

(五) 支撑动态调整存在缺陷

作为“双一流”建设遴选评价和成效评价的重要组成部分,第三方评价在支撑建设动态调整方面仍存在一定缺陷。

一是排名评价在支撑“双一流”建设动态调整运行上存在明显缺陷。排名评价是第三方评价中最具社会影响力的评价模式。以建设资格的动态调整为例,最为朴素的理解便是建立起公开、统一的评价标准,在每个建设周期期末时,对包括在建和候选的高校、学科进行评价并排序,依据排名结果将排在建设总量范围内的高校与学科纳入至新一轮的“双一流”建设名单中。显然,这种评价模式在支撑“双一流”建设动态调整运行中存在缺陷:一方面,大学与学科多样性决定了科学合理的统一评价标准并不存在;另一方面,大学/学科排名中各高校的得分都是其与排名第一高校比较的结果,整体呈现出“排名越是靠后,得分区分度越低,排名结果失真度越大”的特征,依据排名结果建立“双一流”建设动态调整运行机制,部分“双一流”建设高校与部分“双非”高校就会以极其微弱的得分表现互换“身份”^{[12][14]},加之排名评价本就科学性和公正性上存在巨大争议,或将导致“双一流”建设动态调整陷入巨大争议之中。

事实上,“双一流”建设的建设资格动态调整在“进”与“出”上分别依据遴选评价与成效评价等两套评价体系。在“进”的方面,候选高校要想在第一轮建设周期获得建设学科资格,须在“国内权威学科评估结果”“国际上有关大学学科评价结果”“做出杰出研究贡献(国家高等科技奖和高规格的政府人文社科奖)”“区域特色、民族特色与新学科”等“四渠道”中满足任何一条设定标准^[5]。在“出”的方面,建设高校能否继续获得下一轮建设资格主要依据其在建设周期期末的成效评价结果,该成效评价结果包含高校自我评价、专家评价和第三方评价结果等多个方面,以多维多样化形式呈现,不计算总分、不发布排名。由于“进”与“出”遵循两套不同的评价策略,或将导致“双一流”建设动态调整陷入“只增不减”的境地,严重挑战“双一流”建设启动之初所设定的“总量控制、有进有出”等动态调整原则。

二是第三方评价自身存在的问题以及“双一流”建设评价直接采信第三方评价结果的做法不利于“双一流”建设动态调整机制的稳健运行。鉴于第三方评价在指标设计、数据获取及评价方法上仍面临

科学性、公正性挑战,以何种评价标准或采信何种第三方评价结果作为遴选评价依据有待进一步考究。如仅以获得相关重要奖项作为“双一流”建设资格的入场券,而不对其在学科的发展基础、发展潜力作进一步考察,或将导致被遴选中的学科难以在建设取得预期成效。此外,在遴选评价和成效评价中直接采信第三方评价结果的做法,或将诱发部分高校利用第三方评价规则采取急功近利的做法,破坏高等教育系统的学术生态,不利于“双一流”建设战略目标实现。

四、推动第三方评价更好地服务和促进“双一流”建设

如前所述,目前的第三方评价在监测“双一流”建设成效、推动“双一流”建设发展、服务“双一流”动态调整中存在明显短板与不足。由于这些第三方评价自诞生之日起并不是以服务“双一流”建设为导向的,且这些第三方评价已经在国际或国内形成了广泛的认可度和影响力,因而不可能指望能够对这些第三方评价(无论是境外的还是境内的)进行“改造”,使其能够自主承载起“双一流”建设评价所要承担的重要使命。对此,本文提出从“存量”和“增量”两个层面推动第三方评价更好地服务和促进“双一流”建设,即一方面要在“双一流”建设中正确辨别和规范使用现有第三方评价,另一方面要鼓励发展能够符合“双一流”建设导向的新型第三方评价。

(一) 理性看待世界大学/学科排名的价值,辩证认识我国大学/学科表现,树立正确发展观

上海软科、QS、THES、U. S. News、科睿唯安等世界大学/学科排名评价机构都是商业组织,其开展世界大学/学科排名评价的目的本质上是要为其开展相关经济活动而服务,其评价活动遵循的是市场逻辑而非学术逻辑,其评价结果自然不能作为监测和评价高校“双一流”建设成效的直接依据。但是我们仍需客观地认识到世界大学/学科排名的价值。一方面,这些世界大学/学科排名具有较为广泛的全球影响力,是各国留学生选择留学目的地和目标院校的重要参考依据。对一个国家而言,排在这些世界大学/学科排名前列的大学数量越多,对国家吸引全球高端人才越有利,也能在抑制国内高端人才流失方面发挥作用。另一方面,尽管世界大学/学科排名评价体系存在诸多不合理之处,但不可否认的是,很多世界公认的一流名校都居于这些世界大学/学科排名前列。事实上,分指标考察我国大学在世界

大学/学科排名体系中的表现并与世界名校展开对比,对于认识我国大学与世界一流大学的差距与不足仍具有重要参考意义。

从推进“双一流”建设角度来看,一方面要对我国大学在世界大学/学科排名中的表现保持清醒的认知:我国大学在其中的表现既存在贬低的可能,也存在拔高的可能。前者主要源于英语语言环境下论文发表和声誉调查等评价方法对我国大学不利,后者主要源于受发展基础薄弱和排名利益驱使影响,我国很多大学的排名位次提升主要得益于其在量的指标表现上取得显著进步。此外,代表世界一流大学 and 世界一流学科内涵的评价维度无法体现在评价指标体系中,代表质的评价指标表现也可以“以量取胜”,重视当前贡献而忽视历史贡献等等,都将导致我国大学的真实位次与排名位次间存在偏差。另一方面要对我国大学如何在世界大学/学科排名中提升位次树立正确的发展观:是主要依靠提升量的指标表现还是主要依靠提升质的指标表现?前者如国际师生数、论文总量、国际合作论文占比等指标的数值提升在实践中相对容易实现且能在较短时间内取得效果,后者则要依靠潜心研究真问题、做真研究、产出有国际影响力的成果,自然而然地获得顶级期发文机会、提高论文引用次数、提升学术声誉。近年来,我国大学与学科入选世界大学/学科排名的数量与位次都得到显著增加与提升。但不容忽视的是,我国排在世界大学排名前列的大学在表征科研规模的指标得分远超相应排名位次大学的均值,在表征卓越科研的指标得分上却相对落后。如果说以往我国大学发展是要先解决从无到有的问题,那么现在是时候聚焦在内涵式发展的道路上了。

(二)建设健康、安全的第三方评价生态,促进第三方评价的规范化、多元化、自主化发展

当前,很多第三方评价不能承受其重,但符合“双一流”建设导向的第三方评价又供给不足。对此,应从建设第三方评价生态出发,促进符合“双一流”建设导向的第三方评价持续涌现。

一是运用元评估手段规范第三方评价行为。一方面要通过设立元评估评价机构,制定元评估制度,鼓励开展元评估认证活动,规范国内第三方评价行为,确保评价工作的专业性、科学性和公正性。尤其要对纳入“双一流”建设评价的第三方评价展开元评估,确保采信的第三方评价结果能够保证科学性与公正性,进而引导第三方评价规范相关评价行为,营造健康的第三方评价生态。另一方面要加大元评估理论及相关研究成果的宣传与推广,促进社会公众

能够从评价组织的专业性、评价标准的科学性、评价过程的公正性等维度分析第三方评价行为,理性看待第三方评价结果,使一些不规范的第三方评价能够自我改良或逐渐丧失生存空间。

二是鼓励开展更加多元化的第三方评价活动。当前,我国高等教育第三方评价多为针对大学、学科整体发展水平的排名评价以及一些科研奖项的评选,针对大学内部的专业、课程、教学、就业、管理等内容开展的第三方评价较少。此外,相较国外高等教育评价民间组织在各类评价中普遍采取定量评价与定性评价相结合的评价方法,并注重开展满意度调查与声誉调查,我国高等教育评价民间组织在评价数据获取上局限于从公开渠道获取定量指标数据,即在评价活动中不生产数据,而只是做数据的“搬运工”,且完全依赖“数数字”的评价手段。伴随着信息公开力度的加大,建立一个多维度、多指标、客观数据可溯源的办学信息监测平台,不仅更能加深政府、社会、公众对高等教育发展的了解,并且会对没有核心数据优势的排名评价进行有效替代。未来,应鼓励评价机构、学科学会、专业指导委员会等组织围绕专业、课程、教学、就业、管理等内容开展专门评价和服务,围绕学术声誉、学生满意度、雇主满意度等展开调查,助力大学与学科的内涵式发展。

三是保障第三方评价在数据使用上的自主性。当前,我国很多第三方评价在科研文献数据使用上都严重依赖于国外机构提供的数据库,甚至是直接采信这些机构基于相关数据库二次开发获得的评价结果。如国内相关大学/学科排名机构所采用“高被引科学家”“国际论文”“ESI 学科数”“高被引论文数”等指标的数据获取都有赖于科睿唯安公司的 Web of Science 数据库和基于该数据库开发的 ESI 分析工具。这些数据库及其分析工具的开放使用与否、分析工具的算法设计科学性以及数据处理过程的规范性,都将直接决定了相关评价工作能否继续开展的命运以及评价结果的准确性。为此,一方面要自主建立涉及论文遴选范围、期刊等级、高被引论文作者、期刊—学科归属等内容认定的标准体系,基于国外数据库提供的源数据获取相关评价结果,而不是直接采信国外机构提供的评价结论;另一方面要加强国内期刊的国际化建设,自主建设面向全球科研文献的数据库,在国际竞争中逐步建立中国特色的科研评价标准。

(三)以激发高校与学科的内生动力为导向,加强监测评估体系建设,积极推广学科国际评估

高等教育第三方评价大体可以划分为“符号化

评价”与“真实性评价”两类:前者以排名评价为代表,在当前的绩效主义时代中具有坚实的合法性基础,并在回应外部绩效问责、彰显宏观视野和战略定位,引导“双一流”建设方面具有核心优势;后者以高校自主开展的学科国际评估等为代表,对办学主体提升“双一流”建设成效能够发挥更直接、更大的效用^[19]。目前,在我国高等教育第三方评价中,以排名评价为代表的“符号性评价”占据主导地位,且具有广泛的社会影响力。但必须认识到,排名评价以结果为导向,强调竞争性,本质上是一场“零和游戏”。排名评价虽有评价指标,但无评价标准,它让各具特色的大学/学科在统一框架下进行无休止地比较,呈现的评价结果却只是一些苍白的数字。人们很难从排名评价结果中查找问题,揭示原因,实施改进,也就无法切实发挥评价工作的诊断功能与改进功能。

从推动“双一流”建设角度来看,一方面要认识到大学与学科在“双一流”建设中的主体地位,“双一流”建设目标的实现有赖于大学与学科的自觉性与自主性。因此要积极推动办学主体营造质量文化,鼓励其在内部积极开展诊断式评价。具体地,可以自愿参与、经费支持、不问结果为推动策略,鼓励建设高校自主展开学科国际评估,加快世界水准的“双一流”建设步伐。另一方面要对办学主体的外在表现进行多维监测,以回应政府和社会的绩效问责和激发办学主体的质量意识。具体地,可借鉴欧盟“全球多维大学排名”的做法,由官方委托或资助成立公益性组织,建立多维度、多指标、客观数据可溯源的办学信息监测平台并向社会开放,政府及社会公众可根据需要自主选择相关指标进行评价与排序。

(四)以实现“双一流”建设战略目标为导向,推动动态调整机制向“增量式,多维度”方向变革

如前所述,“进”与“出”采用不同评价策略的做法或将导致“双一流”建设动态调整陷入“只增不减”的境地,但同时也要认识到这种做法的积极意义。现有政策体系下,遴选评价以未入选“双一流”建设范围的学科及高校为评价对象,将具有良好发展基础的学科与高校遴选出来纳入建设范围内进行重点建设;成效评价则以已入选建设范围的学科及高校为评价对象,遵循以评促建原则,通过形成监测、改进与评价“三位一体”评价模式,督促高校落实建设主体责任,治本纠偏,持续提高建设水平。一般地,获得建设资格的学科往往在整体上或某一方向上具备了建成世界一流学科的发展基础,其发展优势会在外部资源支持下进一步凸显,其被原本发展基础

就较弱的学科比选下去的可能性也会很低。因此,遴选评价与成效评价采用不同评价策略的做法,既有利于吸引更多学科和高校纳入“双一流”建设范围,又有利于为建设高校与学科提供宽松环境,使其专注于“双一流”建设事业,这对提升我国高等教育系统整体发展水平、实现高等教育强国建设目标具有重要意义。毕竟,动态调整只是促进“双一流”建设的政策工具,而非“双一流”建设的政策目标^[20]。此外,还应以实现“双一流”建设战略目标为导向,进一步创新动态调整机制。

一是要推动“双一流”建设动态调整机制向“增量式”转变。伴随着我国经济体量的日益扩大,我国将逐步具有支持更多大学与学科高质量发展的经济基础。同时,经济社会的高质量发展也需要更多高水平的大学与学科供给更多优秀的高端人才和科研成果。当前,我国高等教育发展水平也正在整体提升,具备建成世界一流学科条件的学科数量将不断增多,面对国际激烈的竞争形势,应当保持开放态度对其进行支持,以提升建设水平,增强国际竞争力。

二是要优化“双一流”建设“有进有出”的动态调整机制。就遴选评价而言,以获得重要科研奖项作为衡量标准时,还应考察其所在学科的发展基础与潜力,以保障遴选学科在“双一流”建设中能够取得预期成效,避免“因人设岗”和诱发人才存量上的恶性竞争。就成效评价而言,由于现有政策对建设资格取消标准的表述模糊,使得部分建设高校与学科“重遴选、轻建设”。对此,一方面要积极营造质量文化,提升自我质量意识,以自愿参与、经费支持、不问结果为推动策略,鼓励“双一流”建设高校自主展开学科国际评估;另一方面,可借鉴美国大学协会(AAU)依据关键评价指标表现接收会员加入和中止会员资格的做法,以卓越为导向构建核心评价指标体系,综合考虑历史贡献与当前贡献(分别赋予不同权重)^{[12]105},对建设高校与学科进行周期性资格审查。

三是要将“双一流”建设的动态调整机制推向更多新维度。“双一流”建设的动态调整并不局限于建设资格的新增与取消,还涉及建设经费的加大与减少。实践中,应积极发挥动态支持政策对建设高校与学科的激励作用。此外,目前的“双一流”建设总体上是依据国家一级学科目录,自上而下地采取扶优策略实施的。面对日新月异的科技发展形势以及新兴交叉学科的不断涌现,可以考虑结合国家战略需求制定交叉学科建设单列名录,允许“双一流”建设高校内部、“双一流”建设高校间、“双一流”建设高

校联合其他院校、科研院所、企业等竞争获得建设资格,推动交叉学科建设与发展,抢占前沿科技制高点,提升高等教育对国家科技创新发展的贡献度。

参考文献:

- [1] 郭宝宇,冯俾琳.世界一流大学第三方评价机构的识别系统:外部质量保障视角[J].清华大学教育研究,2020,41(6):83-90.
- [2] 李亚东,俎媛媛.我国第三方教育评价的核心问题辨析及政策建议[J].教育发展研究,2018,38(21):1-5.
- [3] 郑佳斯,卜熙.失效的第三方:组织自利性下的社会组织评估[J].华南师范大学学报(社会科学版),2020(5):98-112.
- [4] 曹晶,车丽萍.依附式自主:高等教育评价中第三方组织与政府的关系重构[J].中国高教研究,2021(9):59-64.
- [5] 杨卫.“双一流”建设的上行式评估[J].大学与学科,2020,1(1):145-153.
- [6] 杨卫,龚旗煌,杨斌,等.《第五轮学科评估工作方案》专家解读[J].大学与学科,2021,2(1):117-128.
- [7] 俞巢.大学评估何处去?国际评估在中国一流大学的兴起、扩散与制度化[J].华东师范大学学报(教育科学版),2022,40(1):30-42.
- [8] 周汉华.全面依法治国与第三方评估制度的完善[J].法学研究,2021,43(3):19-35.
- [9] 尚虎平,王春婷.政府绩效评估中“第三方评估”的适用范围与限度——以先行国家为标杆的探索[J].理论探讨,2016(3):12-18.
- [10] 加快构建中国特色“双一流”建设评价体系——教育部学位管理与研究生教育司负责人就“双一流”建设成效评价工作答记者问[EB/OL].http://www.moe.gov.cn/jyb_xwfb/s271/202103/t20210323_521943.html.
- [11] 陈玉琨.教育评价学[M].北京:人民教育出版社,2005:7-10.
- [12] 崔育宝.我国“世界一流大学”建设评价研究[D].合肥:中国科学技术大学,2018:44,141,105.
- [13] 吴剑平,陈星博,孙茂新.一流大学评价的基本问题探讨[J].教育发展研究,2002(12):46-49.
- [14] 张应强.教育内外部关系规律及其在高等教育研究中的运用[J].复旦教育论坛,2020,18(5):5-11.
- [15] 穆蕴秋,江晓原.“影响因子”游戏之父尤金·加菲尔德评传[J].自然辩证法通讯,2017,39(6):145-154.
- [16] 张应强.“双一流”建设需要什么样的学科评估——基于学科评估元评估的思考[J].清华大学教育研究,2019,40(5):11-18.
- [17] 武建鑫,周光礼.世界一流学科:“以评促建”何以可能——基于系统科学的分析[J].国家教育行政学院学报,2016(11):53-61.
- [18] 操太圣.遭遇问责的高等教育绩效化评价:一个反思性讨论[J].南京社会科学,2018(10):129-136.
- [19] 朱冰莹,董维春.从“符号化评价”到“真实性评价”:一流学科建设评价的核心向路[J].学位与研究生教育,2018(6):1-6.
- [20] 张应强.正确认识“双一流”建设成效评价与动态调整的关系[J].西北工业大学学报(社会科学版),2021(3):32-37.

On Third-Party Evaluation in “Double First-Class” Construction: From “Evaluation Worship” to “Evaluation Enabling”

CUI Yubao, LI Jinlong, ZHANG Shulin

(School of Public Affairs, University of Science and Technology of China, Hefei 230026, China)

Abstract: The third-party evaluation, highly recognized for its independence, professionalism and impartiality, has become an important part of the “double first-class” construction. The current third-party evaluation practice has some problems, such as low comparison with world-class quality, weak evaluation guidance to contributions to the society, difficulty in guaranteeing scientificity and impartiality, insufficient role in promoting the “double first-class” construction, and defects in supporting dynamic adjustment. Therefore, it can hardly bear the expected heavy tasks. This paper suggests that in the future, the attitude towards third-party evaluation should be changed from “evaluation worship” to “evaluation enabling” in the process of “double first-class” construction. Specifically, the third-party evaluation should be pushed forward from the “stock” and “increment” aspects so that it can better serve and promote the “double first-class” construction. That is, on the one hand, we should correctly understand and standardize the application of the existing third-party evaluation system, and on the other hand, we should encourage development of new third-party evaluation system that conforms to the orientation of “double first-class” construction.

Keywords: third-party evaluation; “Double First-Class” construction; meta-evaluation